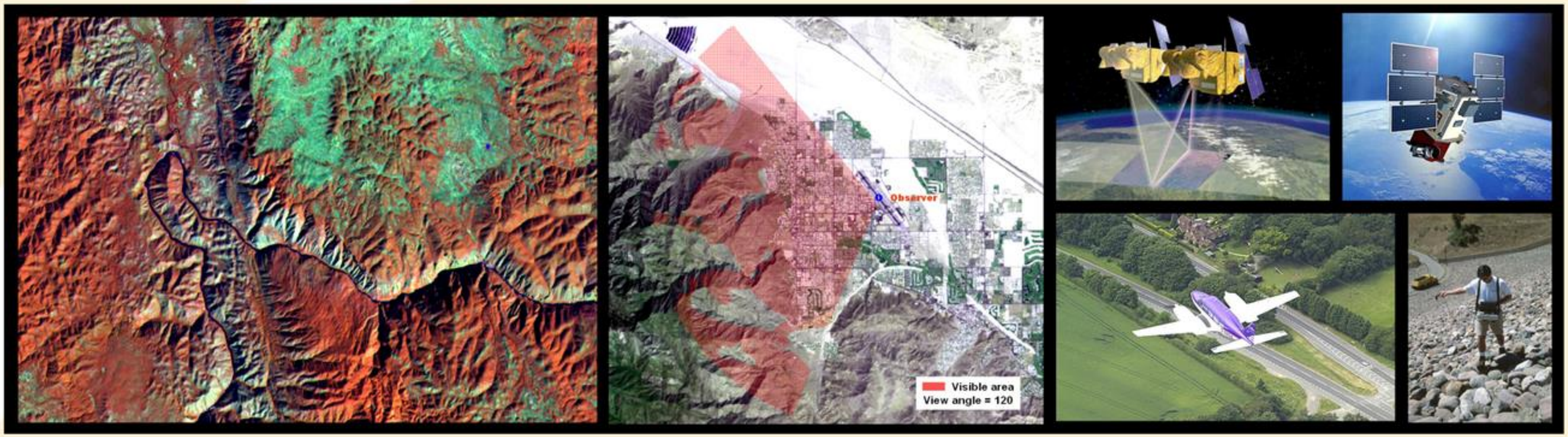
Division of Spatial Information Science
Graduate School of Life and Environmental Sciences
University of Tsukuba

Fundamentals of Remote Sensing and its Applications in GIS



VLM (Visual Learning Material), Read Less, Learn More .....

Ko Ko Lwin
Division of Spatial Information Science
Graduate School of Life and Environmental Sciences
University of Tsukuba

Preface

Remote Sensing data is one of the primary data sources in GIS analysis. The objective of this material is to provide fundamentals of Remote Sensing technology and its applications in Geographical Information Systems to undergraduate students and the one who wants to study about Remote Sensing technology by visually (Read less learn more).

However, Remote Sensing technology had been well established for several decades and still booming. Handling and interpretation of remote sensing data will never be easy. It requires additional practical works and digital image processing knowledge. It is impossible to cover all topics in here. So, here I provide additional learning information and other online resources which were listed in Appendix A for further interested students.

I hope you'll enjoy it

Ko Ko Lwin
Division of Spatial Information Science
University of Tsukuba
2008

Contents

Part I: Fundamentals of Remote Sensing

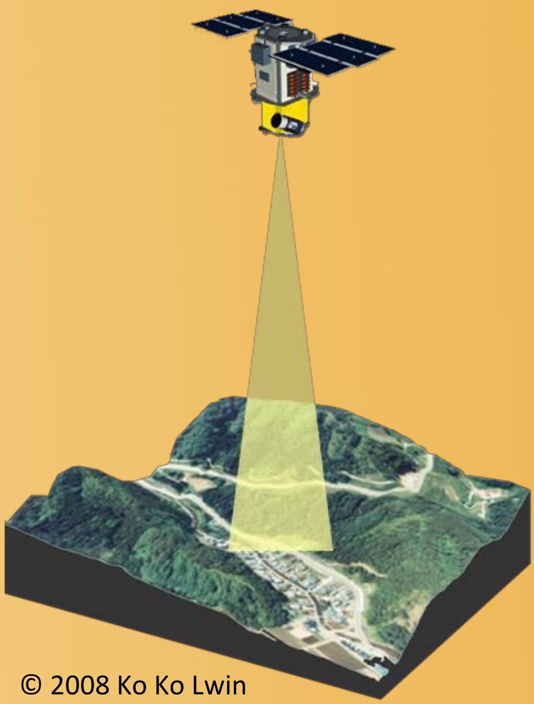
Part I: Fundamentals of Remote Sensing

1. Remote Sensing Overview

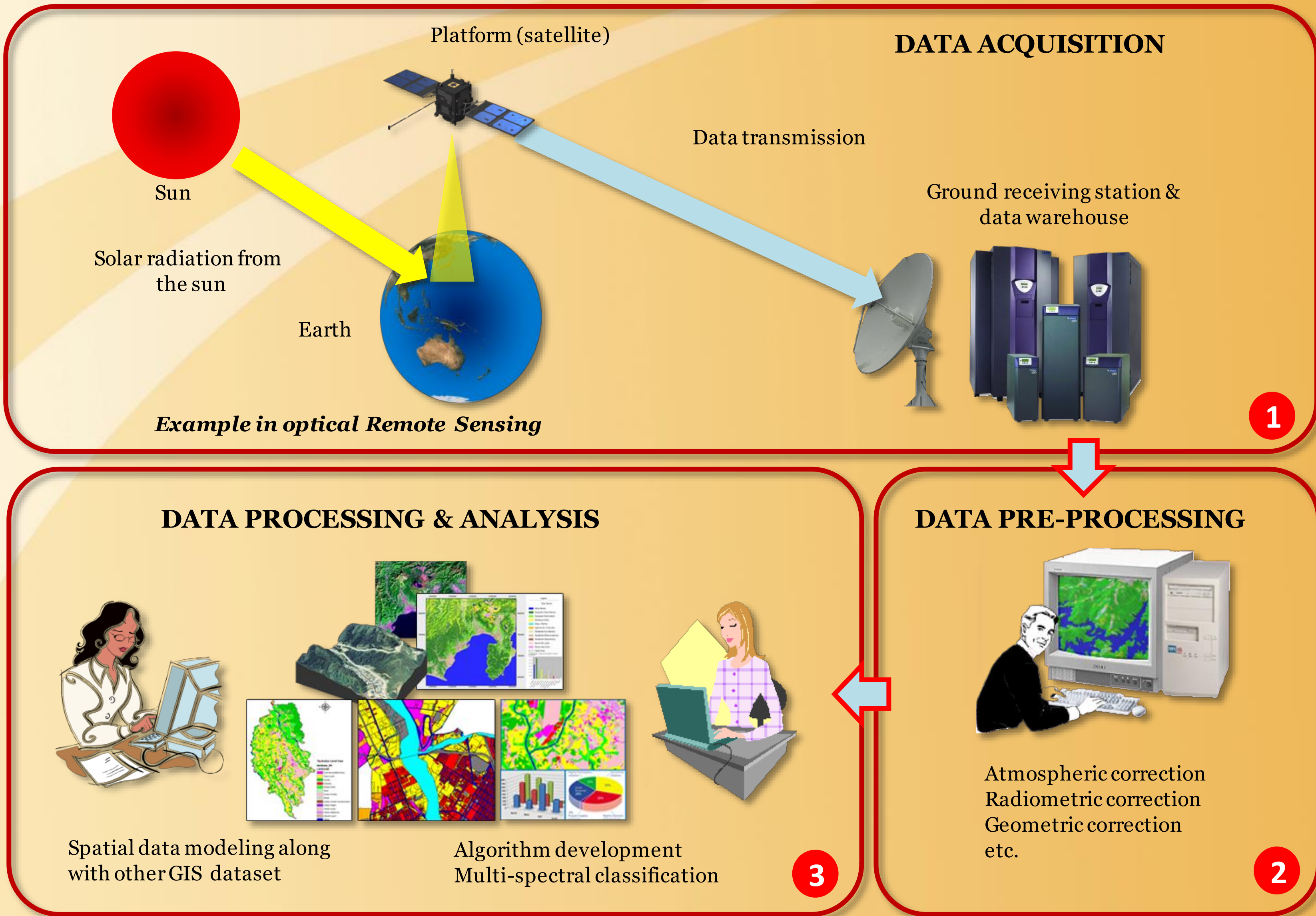
1.1 Definition

What is Remote Sensing? If you are reading this sentence, now you are doing Remote Sensing. In fact, any information acquired from the object without touching is Remote Sensing. Following is a scientific definition of Remote Sensing.

The **science of acquiring information about the earth** using instruments which are remote to the earth's surface, usually from aircraft or satellites. Instruments may use visible light, infrared or radar to obtain data. Remote sensing offers the ability to observe and collect data for large areas relatively quickly, and is an important source of data for GIS. (Source: digimap)



1.2 Remote Sensing and GIS Work Flow



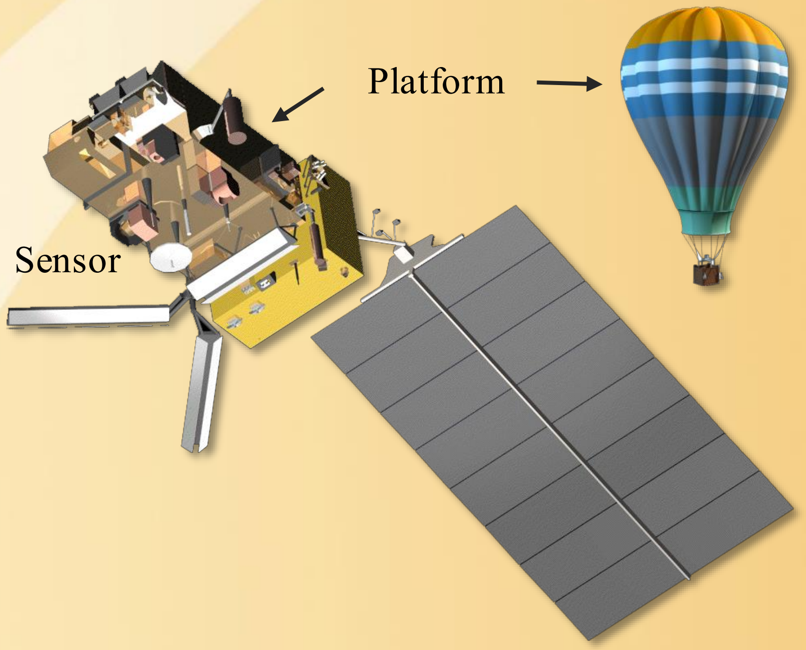
1.3 Components in Remote Sensing

Platform

The vehicle which carries a sensor. i.e. satellite, aircraft, balloon, etc...

Sensors

Device that receives electromagnetic radiation and converts it into a signal that can be recorded and displayed as either numerical data or an image.



One platform can carry more than one sensor. For example:

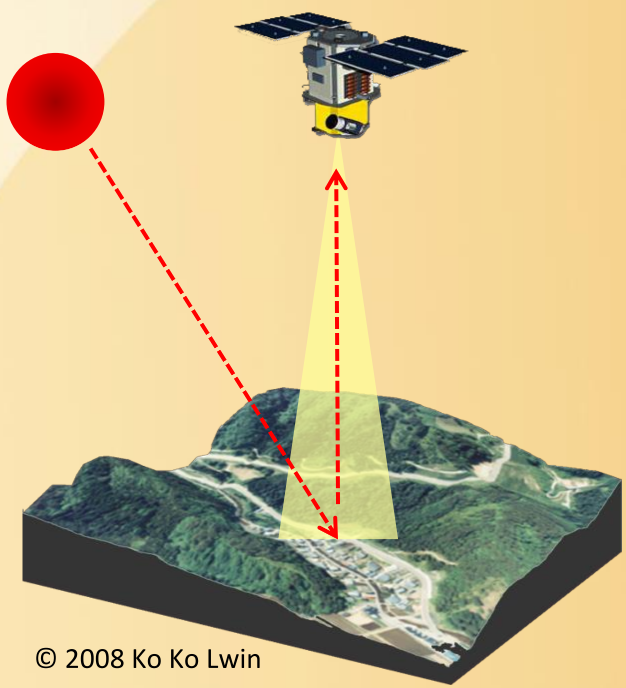
Platform Name	Sensor Name
Landsat TM	Thematic Mapper (Passive: Optical sensor)
Landsat ETM	Enhanced Thematic Mapper (Passive: Optical sensor)
ALOS	PRISM (Passive: Optical sensor) AVNIR-2 (Passive: Optical sensor) PALSAR (Active: Microwave sensor)

1.4 Types of Remote Sensing

Passive Remote Sensing and Active Remote Sensing

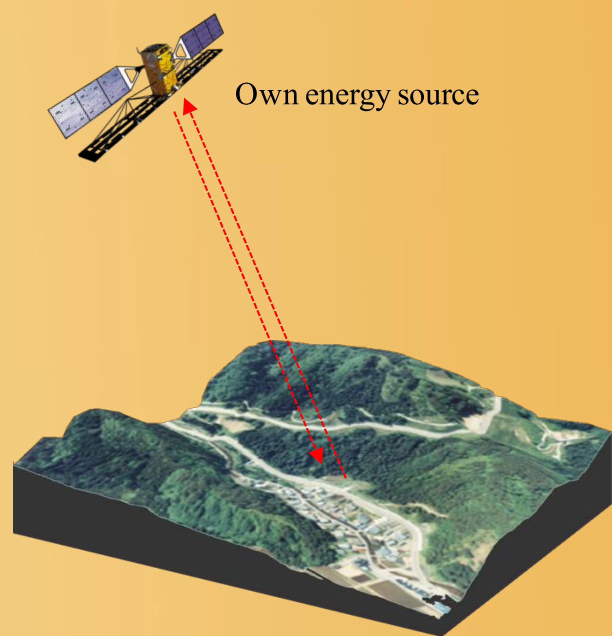
Passive Remote Sensing

Remote sensing of energy naturally reflected or radiated from the terrain.

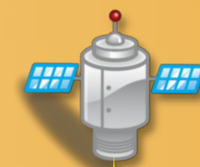


Active Remote Sensing

Remote sensing methods that provide their own source of electromagnetic radiation to illuminate the terrain. Radar is one example.



1.5 Multistage Remote Sensing Data Collection



Satellite based remote sensing

Advantages: Less geometric errors (platform is stable)
Disadvantages: Need to wait a time for certain event
Fixed spatial resolution

Aerial surveying

Advantages: Acquire any times any events
Variable spatial resolution by changing flight altitude and camera focal length
Disadvantages: High geometric errors; require sophisticated geometric correction model
Costly for specific area, specific purpose

Ground based remote sensing GBRS or Low Altitude Remote Sensing

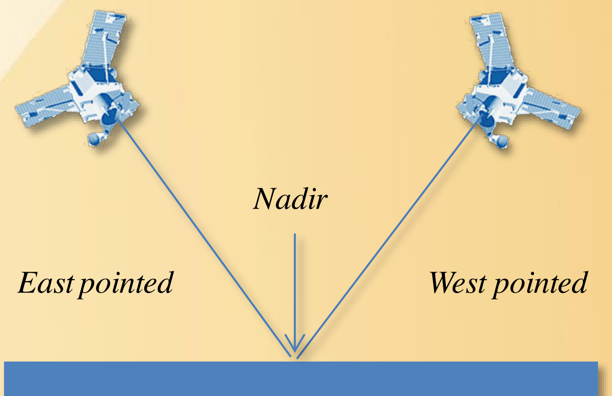
Scientific experiment purposes
(e.g. study about canopy, soil contamination, etc.)

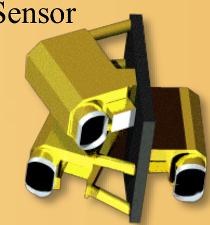
1.6 Stereo Pair Remote Sensing Data Collection

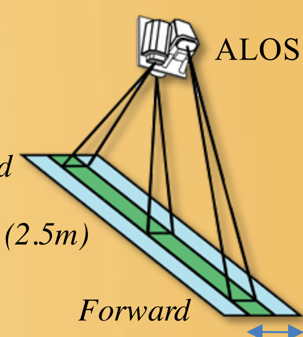
Some satellites capable to acquire stereo pair images that can be achieved when two images of the same area are acquired on different days from different orbits, one taken East of the other (i.e., East or West of the nadir). For this to occur, there must be significant differences in the inclination angles.

Stereo Pair Images

- Different days from different orbits
  e.g. IKONOS's pointing system
- Onboard two or more inclined sensors
  e.g. SPOT5, ALOS PRISM (Panchromatic Remote-sensing Instrument for Stereo Mapping)







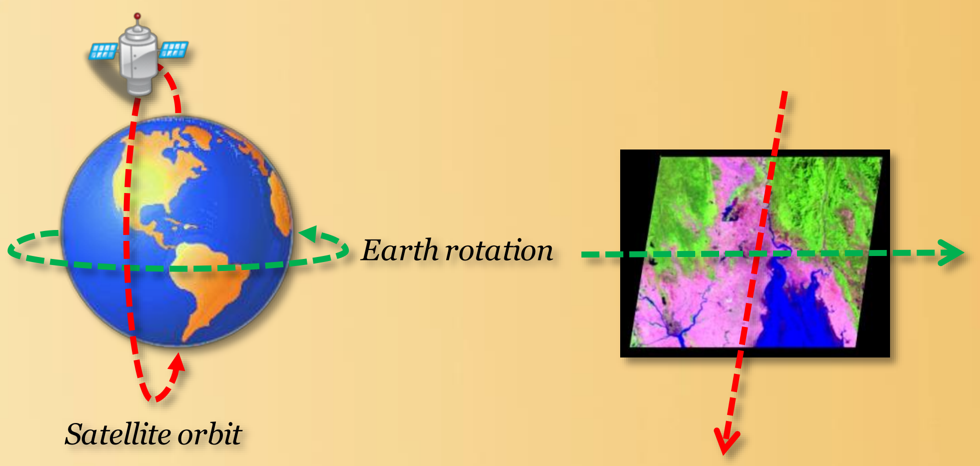
DEM Digital Elevation Model can be generated from stereo pair images by using Photogrammetric technology.

1.7 Types and Uses of Satellites

Types of satellites can be classified by their orbit characteristics.

Type 1: Low Earth Orbits/Satellites: Normally used in spy satellite (Military purposes)

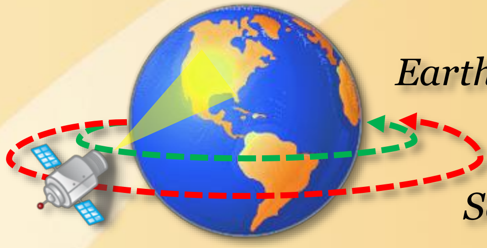
Type 2: Sun-synchronous Orbits/Satellites: a polar orbit where the satellite always crosses the Equator at the same local solar time. Most of the earth resources satellites are sun-synchronous orbit.

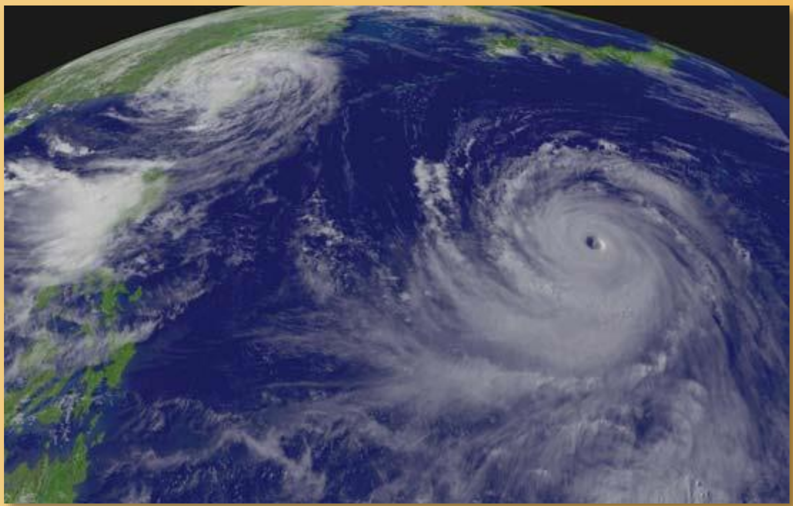


Examples

Landsat TM/ETM
SPOT
ALOS
IKONOS
QuickBird

Type 3: Geostationary Orbits/Satellites: Satellites at very high altitudes, which view the same portion of the Earth's surface at all times. Especially used in metrological applications.

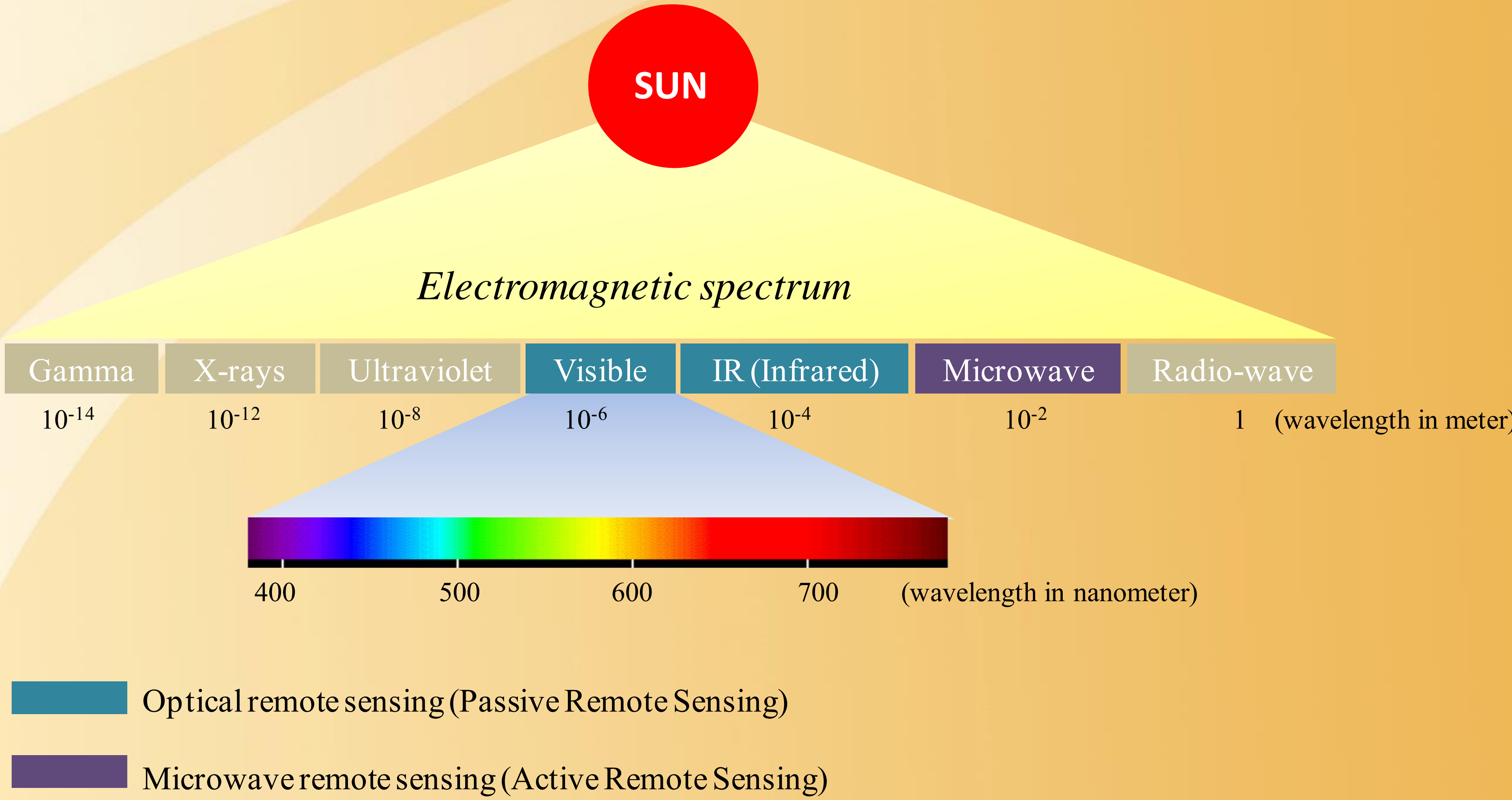




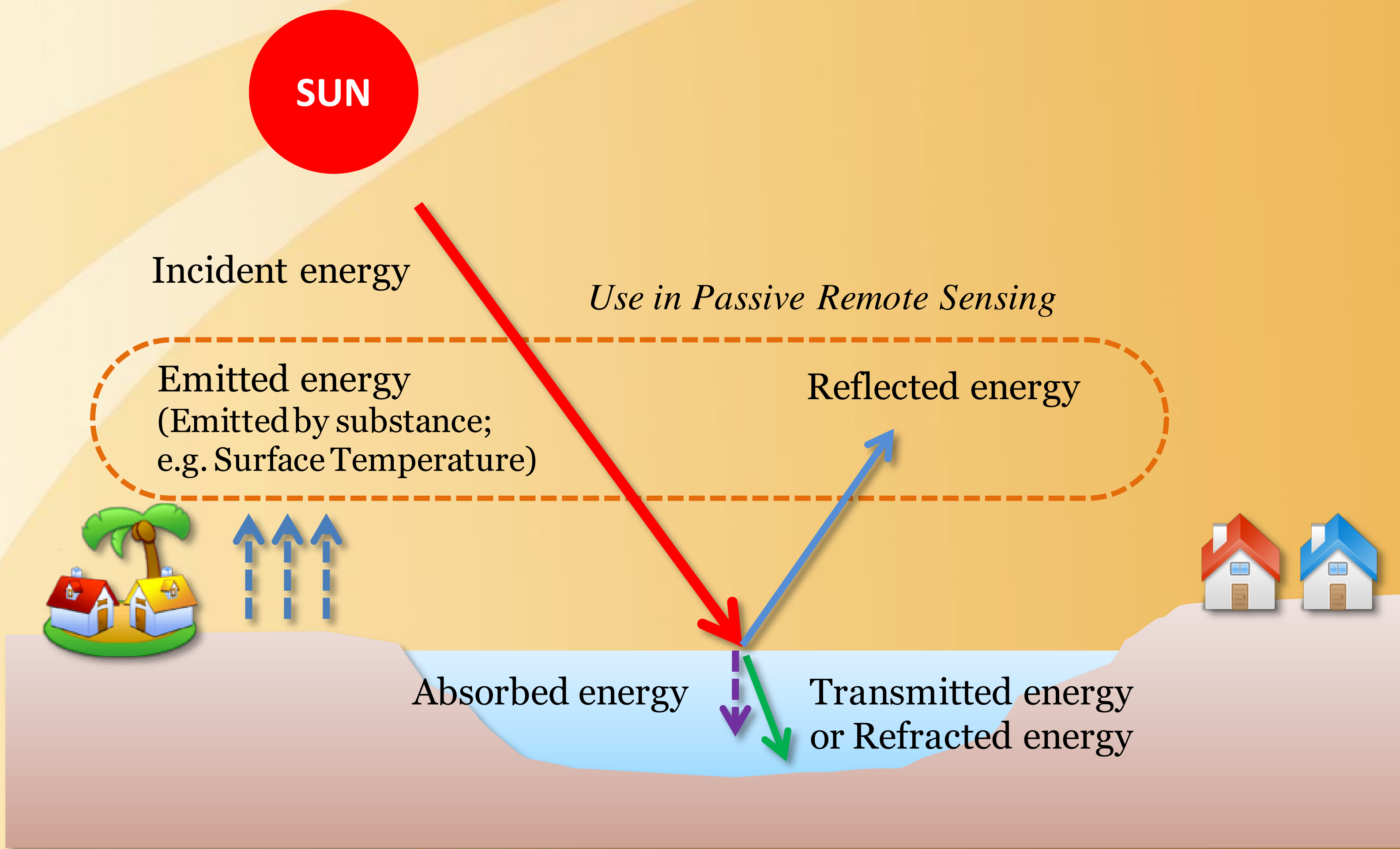
- *Fixed position on specific location*
- *Same speed as earth rotation speed*
- *Wide area coverage*
- *Especially designed for weather monitoring*

2. Remote Sensing Data Acquisition

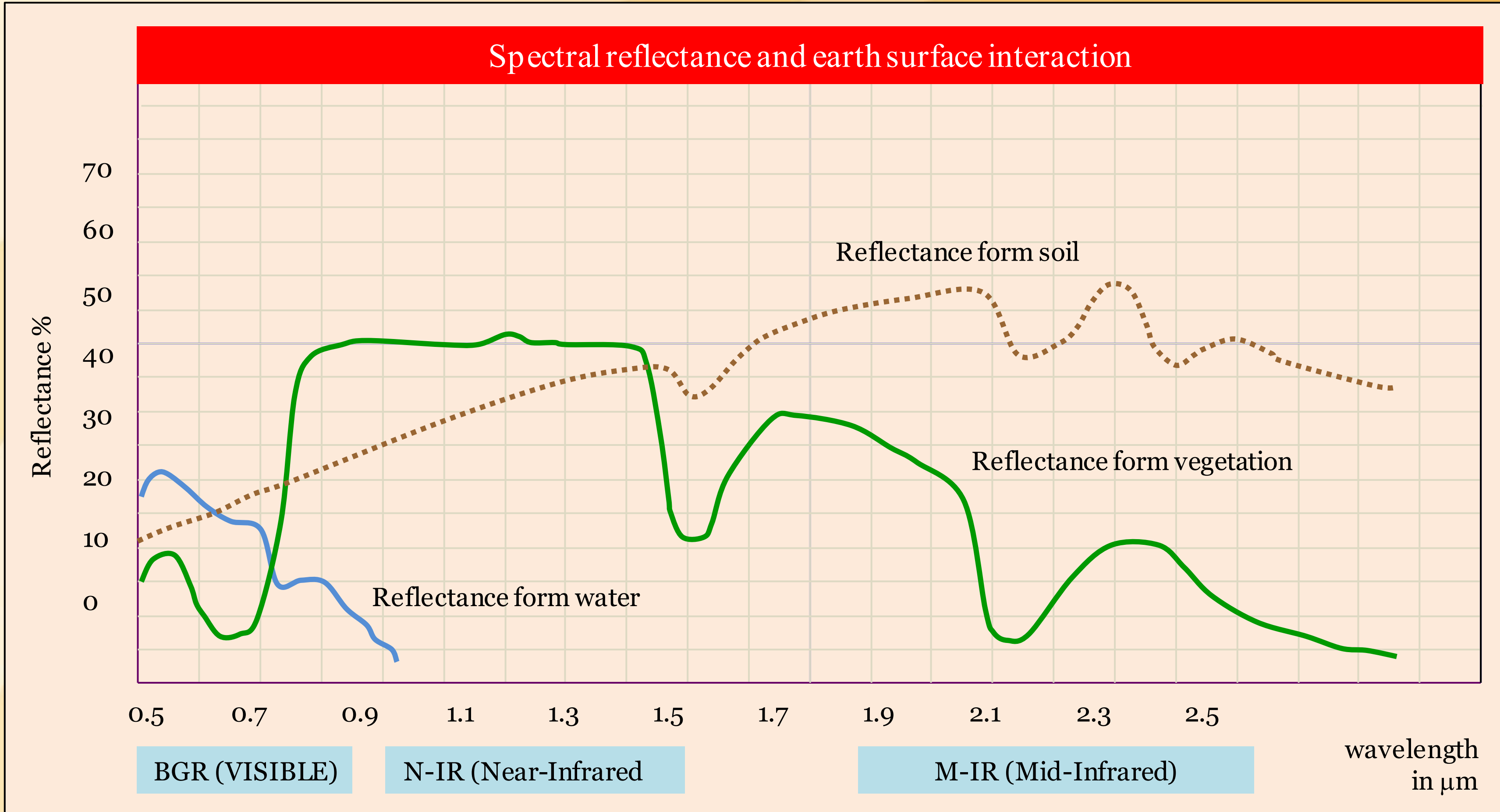
2.1 Electromagnetic Waves Used in Remote Sensing



2.2 Properties of Electromagnetic Waves



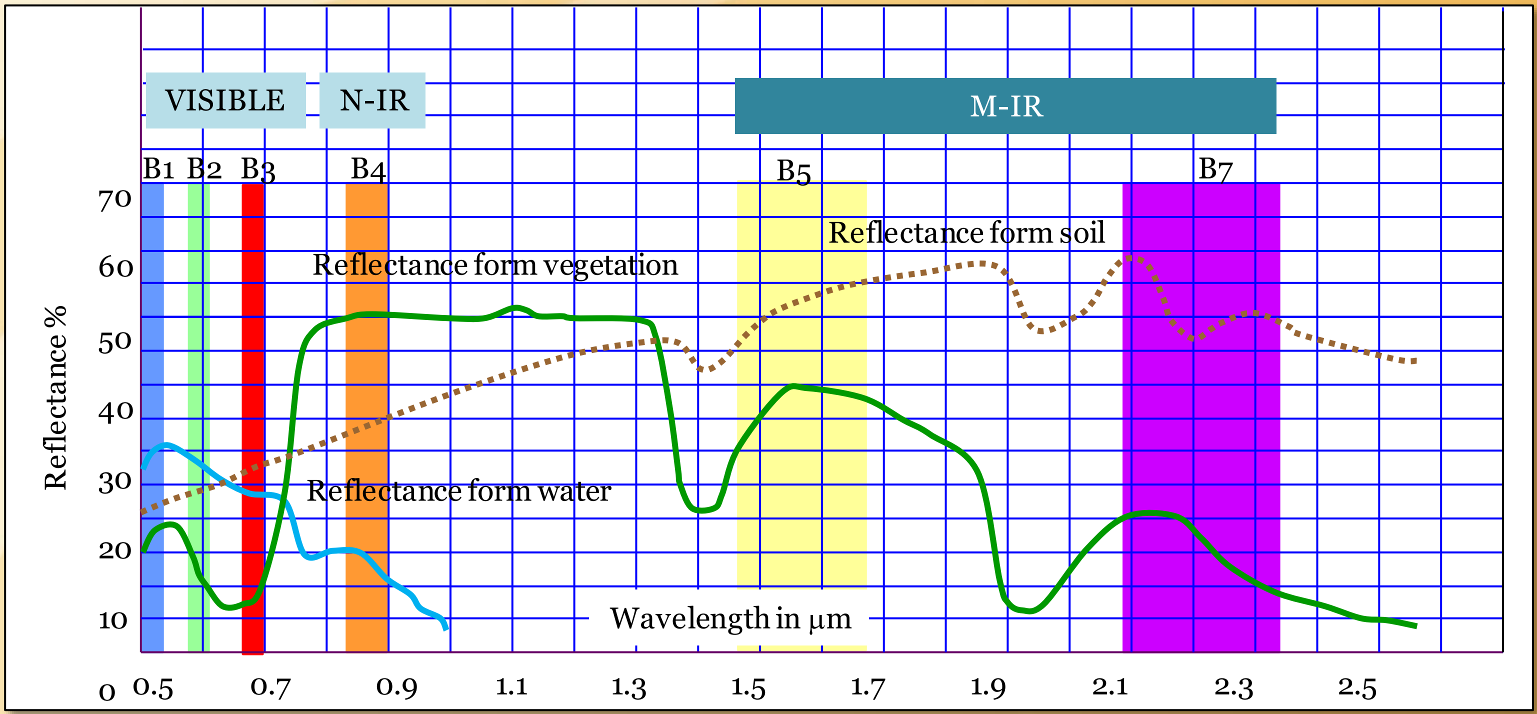
2.3 Spectral Reflectance and Earth Surface Interaction

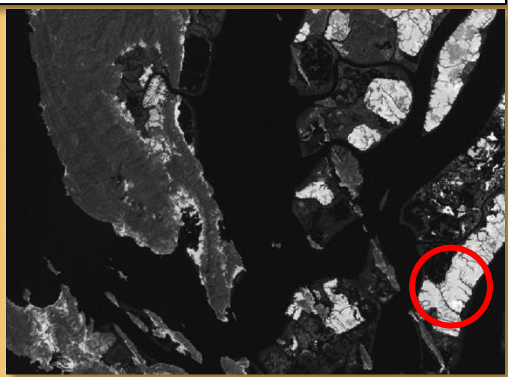


Surface category	Low reflectance	High reflectance
Water	N-IR (Near -Infrared)	Blue (Visible)
Vegetation	M-IR (Mid-Infrared)	N-IR (Near-Infrared)
Soil	Blue (Visible)	M-IR (Mid-Infrared)

2.4 Multi-spectral Remote Sensing Data (Image)

- Composed with more than one spectral band and each band represents specific wavelength
- Example in Landsat TM (Total 7 bands, Band 6 Thermal band omitted in here)



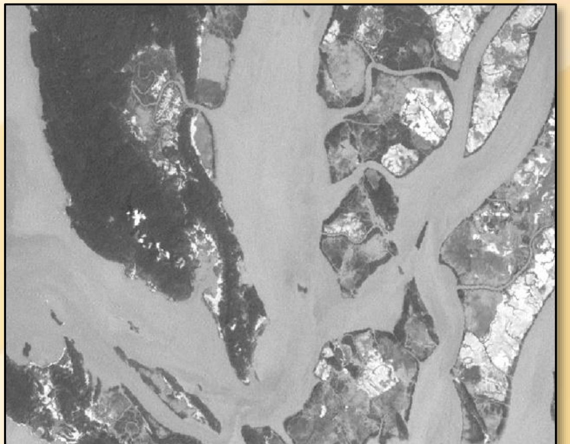


TM Band 1: High reflectance in water **TM Band 4**: High reflectance in vegetation **TM Band 7**: High reflectance in bare land (soil)

2.4 Multi-spectral Remote Sensing Data (Image) (Continued)

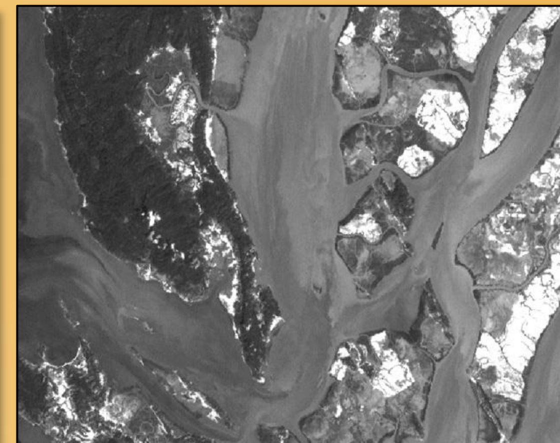
Example in Landsat TM/ETM (Band 6 omitted)

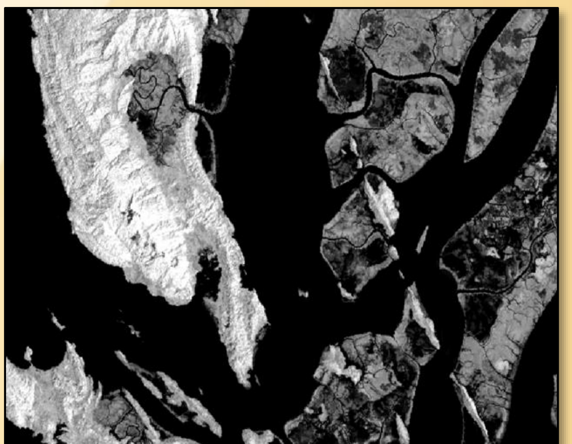
Band 1 : Blue (0.450 ~ 0.515 μm)



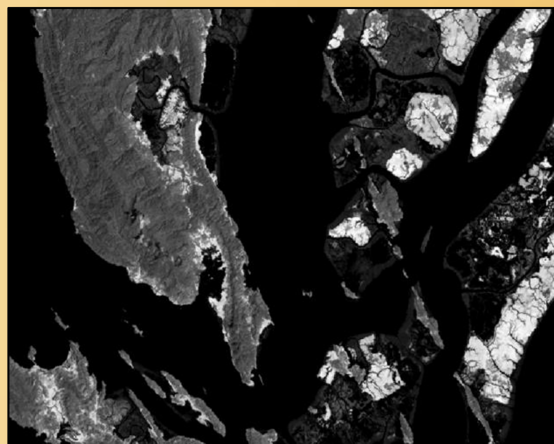
Band 2 : Green (0.525 ~ 0.605 μm)

Band 3 : Red (0.630 ~ 0.690 μm)

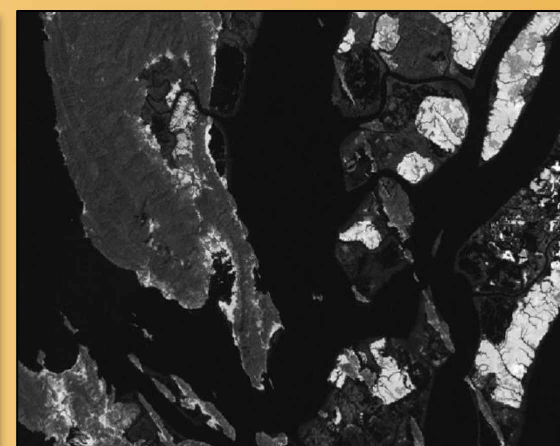




Band 4 : Near-Infrared(0.750 ~ 0.900 μm)



Band 5 : Mid-Infrared (1.550 ~ 1.750 μm)



Band 7 : Mid-Infrared (2.090 ~ 2.350 μm)

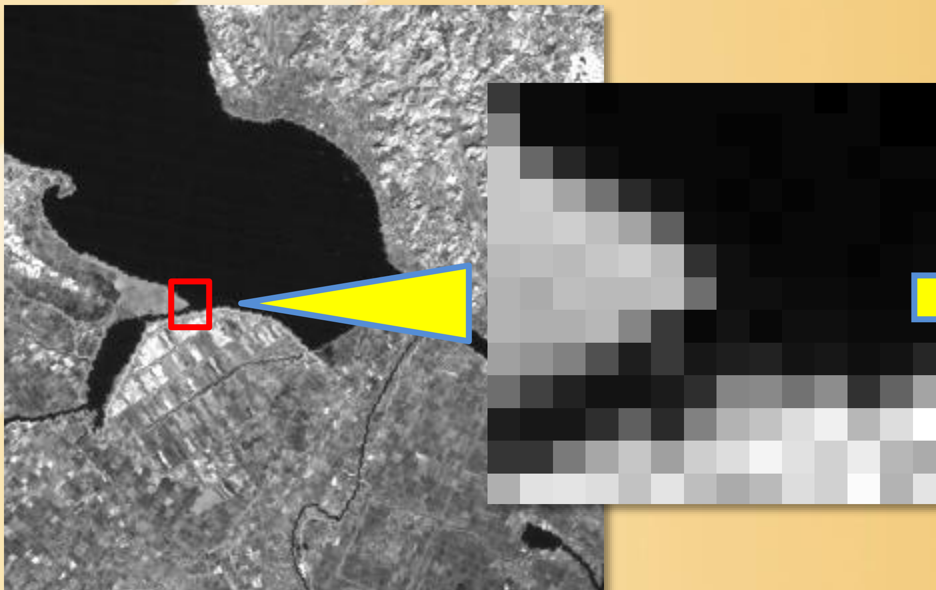
2.5 Spectral Properties and Principal Applications

Example in Landsat TM/ETM

Band	Wavelength (µm)	Principal applications
B-1	0.45 - 0.52 (Blue)	This band is useful for mapping coastal water areas, differentiating between soil and vegetation, forest type mapping, and detecting cultural features.
B-2	0.52 - 0.60 (Green)	This band corresponds to the green reflectance of healthy vegetation. Also useful for cultural feature identification.
B-3	0.63 - 0.69 (Red)	This band is useful for discriminating between many plant species. It is also useful for determining soil boundary and geological boundary delineations as well as cultural features.
B-4	0.76 - 0.90 (Near-Infrared)	This band is especially responsive to the amount of vegetation biomass present in a scene. It is useful for crop identification and emphasizes soil/crop and land/water contrasts.
B-5	1.55 - 1.75 (Mid-Infrared)	This band is sensitive to the amount of water in plants, which is useful in crop drought studies and in plant health analyses. This is also one of the few bands that can be used to discriminate between clouds, snow, and ice.
B-6	10.4 - 12.5 (Thermal Infrared)	This band is useful for vegetation and crop stress detection, heat intensity, insecticide applications, and for locating thermal pollution. It can also be used to locate geothermal activity.
B-7	2.08 - 2.35 (Mid-Infrared)	This band is important for the discrimination of geologic rock type and soil boundaries, as well as soil and vegetation moisture content.

2.6 Spectral Reflectance to DN (Digital Number)

In fact, remote sensing data is converting of spectral reflectance value to digital number (DN) known as a pixel. Each spectral wavelength represents as a single layer in remote sensing data called “Band” or “Channel”. The more bands or channels present, the more spectral properties in remote sensing data.



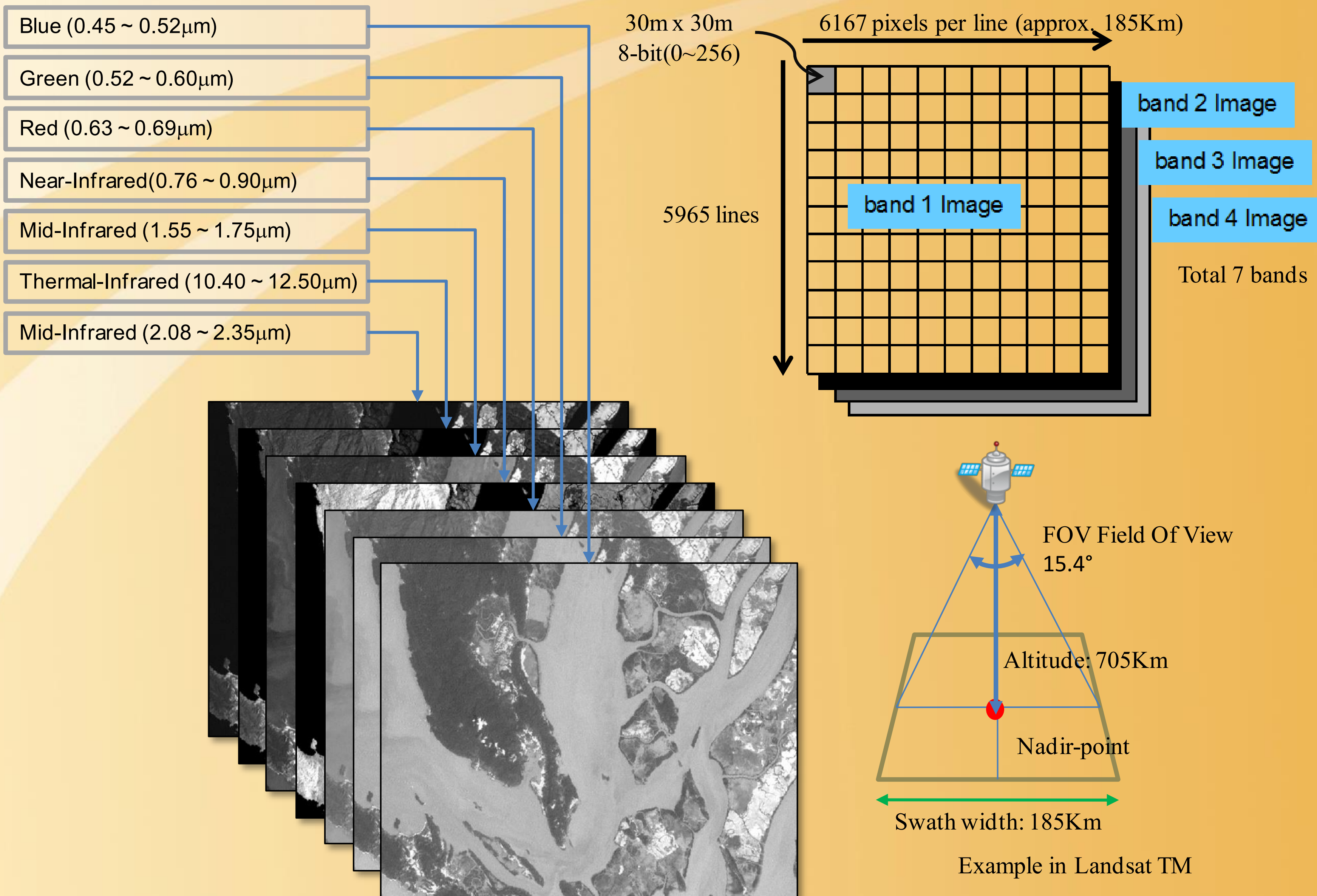
Low DN → Darker

31	19	19	17	18	18	18	18	18	18	16	18	16	16
51	19	19	18	18	18	18	17	17	18	18	18	16	16
68	43	26	20	18	18	18	18	17	18	18	17	18	18
68	69	59	46	27	21	18	17	18	17	18	18	17	17
68	68	70	66	59	41	19	18	17	18	18	18	17	18
65	66	65	68	70	65	29	20	18	18	18	17	18	19
63	61	66	65	65	66	45	20	20	19	19	19	18	19
63	62	62	64	51	31	18	21	18	20	20	19	19	19
58	55	50	37	24	31	22	24	25	21	22	20	21	26
45	33	25	21	21	23	28	51	52	46	53	29	42	59
24	22	22	28	41	27	50	63	67	74	79	64	74	83
30	30	48	60	68	58	70	74	80	75	71	78	64	61
62	75	76	74	67	76	66	61	65	74	71	82	63	76

High DN → Brighter

DN = 0 255

2.7 Structure of Remote Sensing Data *(Example in Landsat TM)*



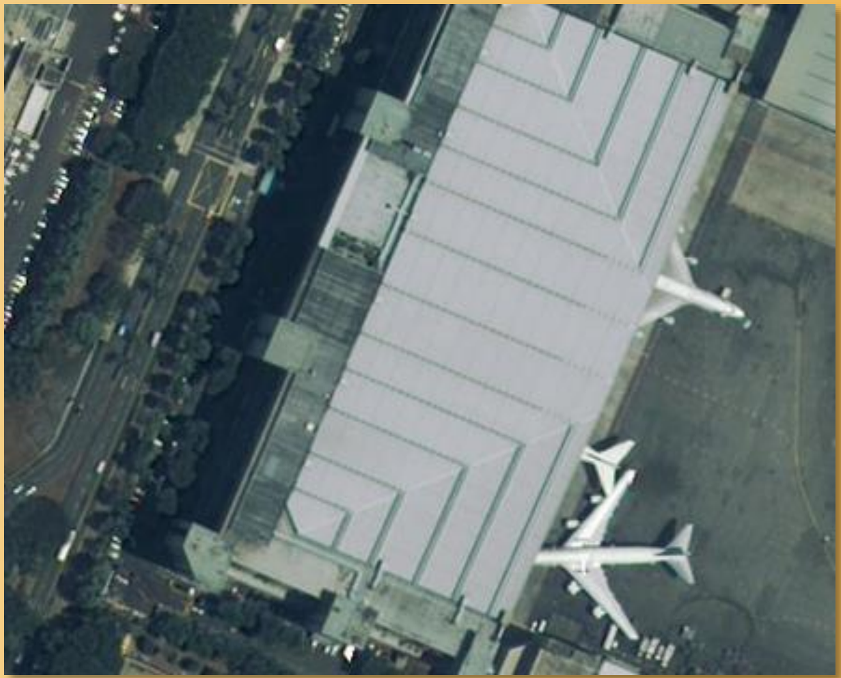
2.8 Resolutions in Remote Sensing

There are four types of resolutions in Remote Sensing.

(a) Spatial Resolution: The detail discernible in an image is dependent on the spatial resolution of the sensor and refers to the size of the smallest possible feature that can be detected.

Example: Landsat TM Spatial resolution 30mx30m, QuickBird 67cm x 67cm

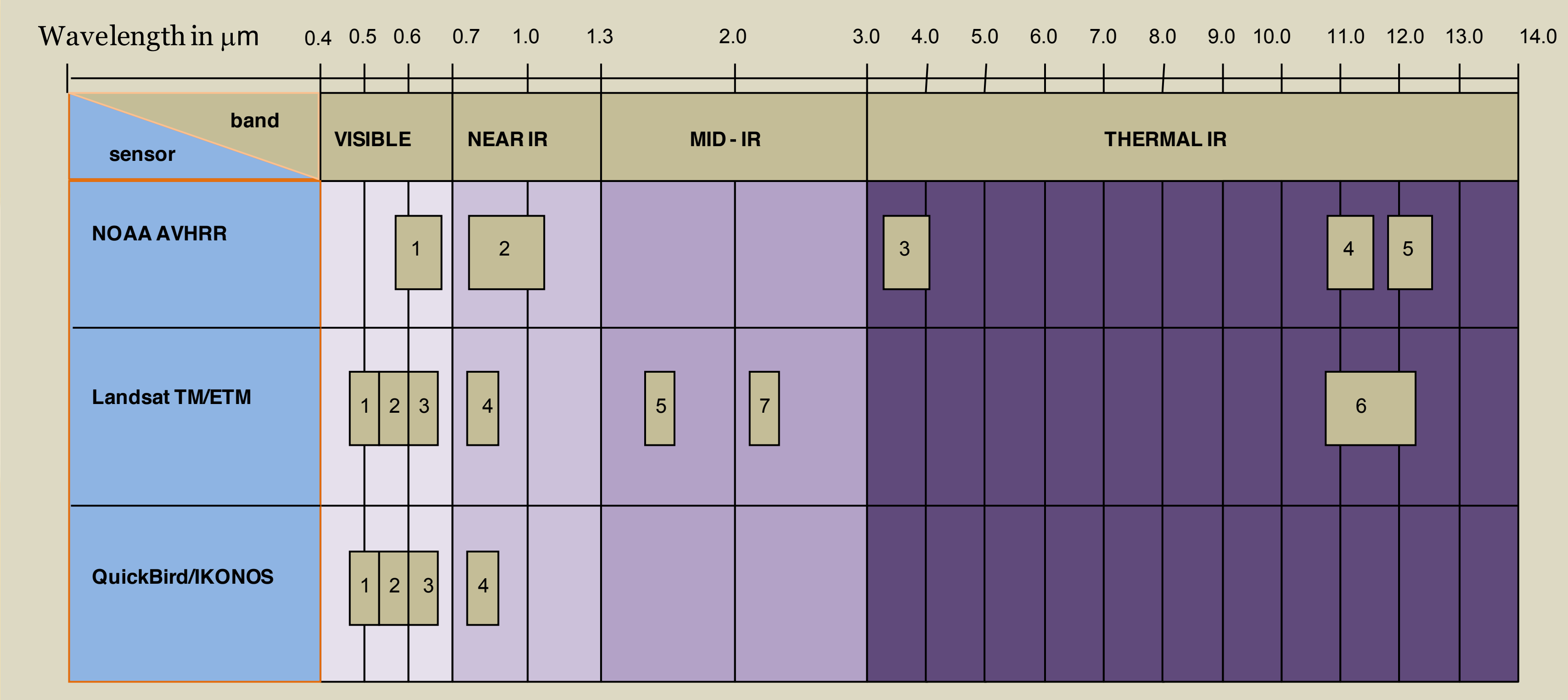
Landsat TM 30m x 30m



QuickBird 67cm x 67cm

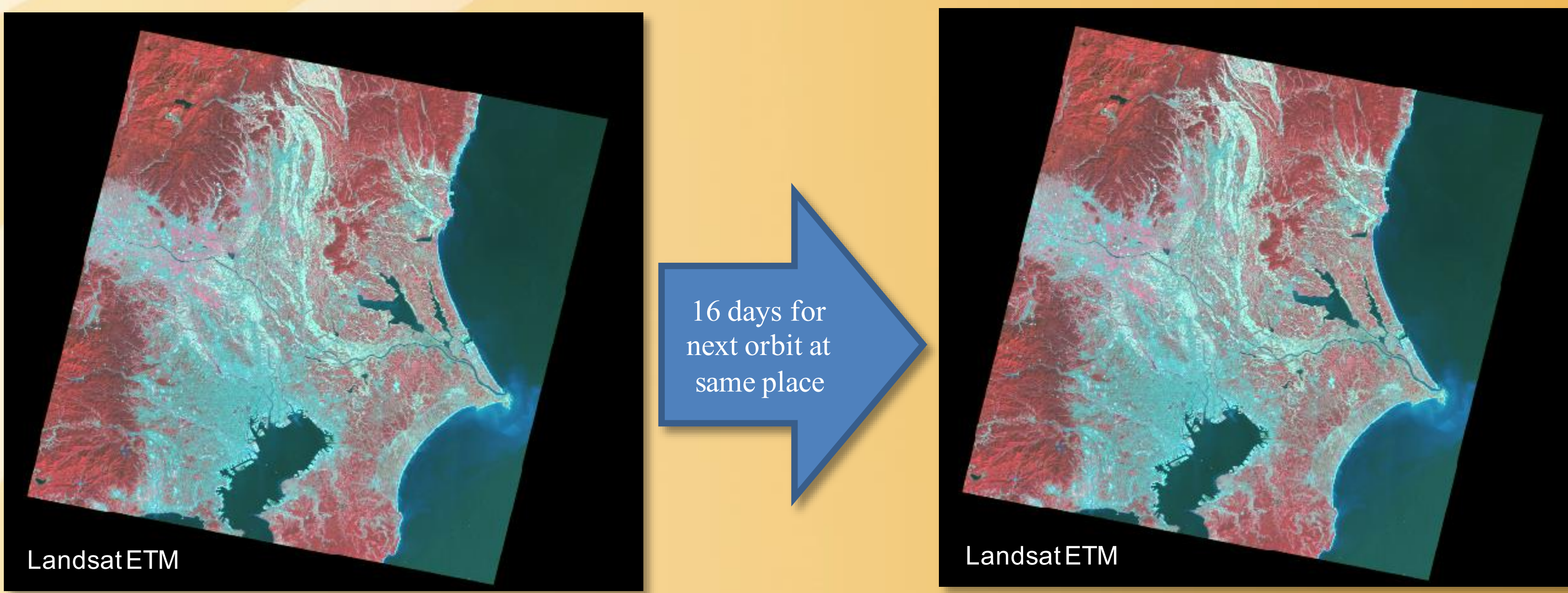
(b) Spectral Resolution: Spectral resolution describes the ability of a sensor to define fine wavelength intervals. The finer the spectral resolution, the narrower the wavelength range for a particular channel or band.

Example: Landsat TM has 7 Bands, QuickBird/IKONOS Multispectral has 4 Bands, etc.



(c) Temporal Resolution: Also important to consider in a remote sensing system, refers to the length of time it takes for a satellite to complete one entire orbit cycle. The revisit period of a satellite sensor is usually several days except Geostationary satellites.

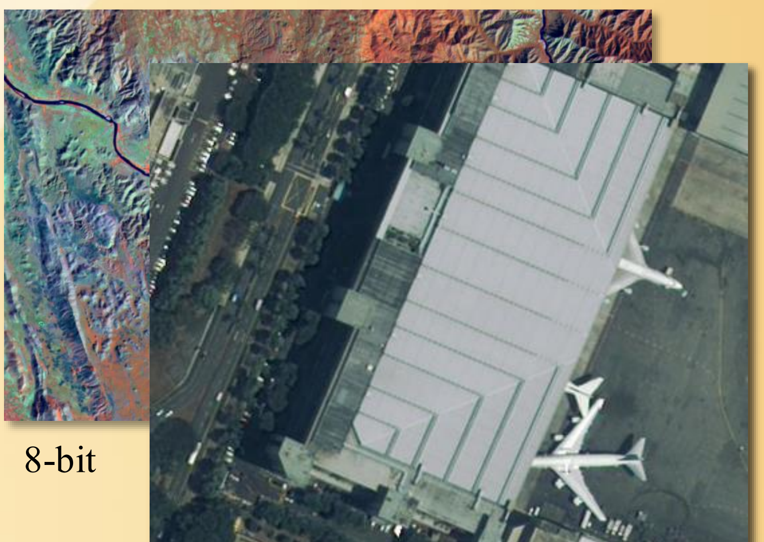
Example: Landsat TM 16 days, SPOT 26 days, etc.



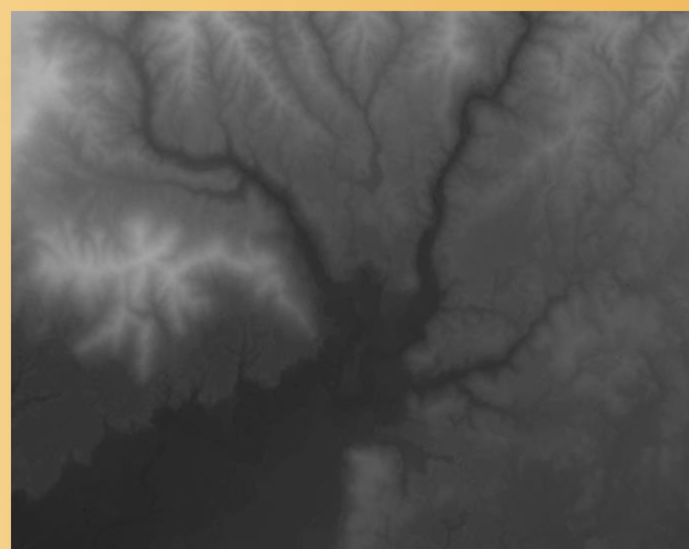
(d) Radiometric Resolution: The radiometric resolution of an imaging system describes its ability to discriminate very slight differences in energy.
The finer the radiometric resolution of a sensor, the more sensitive it is to detecting small differences in reflected or emitted energy.

Example: Landsat TM 8 bits, SPOT 8 bits, IKONOS 11 bits. However, most computer programs do not support 11-bit, so it will convert to 16-bit.

8-bit : 2^8 = maximum 256 color levels or DN values (commonly used)
16-bit : 2^{16} = maximum 65536 color levels or DN values (especially used in elevation data, e.g. DEM, DSM, DTM, etc.)



8-bit



16-bit

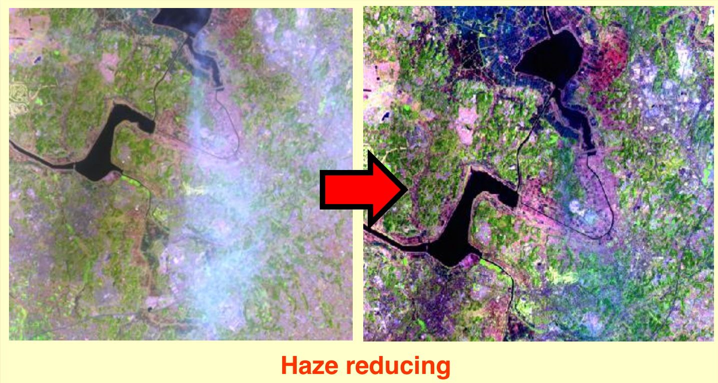
3. Remote Sensing Data Processing and Analysis

3.1 Remote Sensing Data Pre-processing

(a) Atmospheric correction
(b) Radiometric correction
(c) Geometric correction

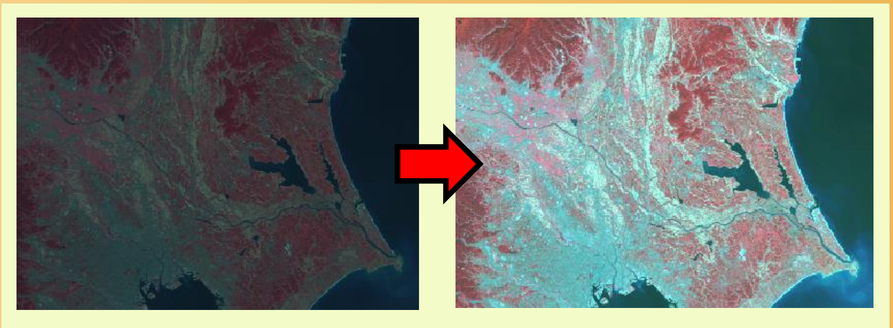
However, most remote sensing data can be acquired or purchased atmospheric, radio metric and geometric corrected data.
Here, we will introduce briefly.

(a) Atmospheric correction

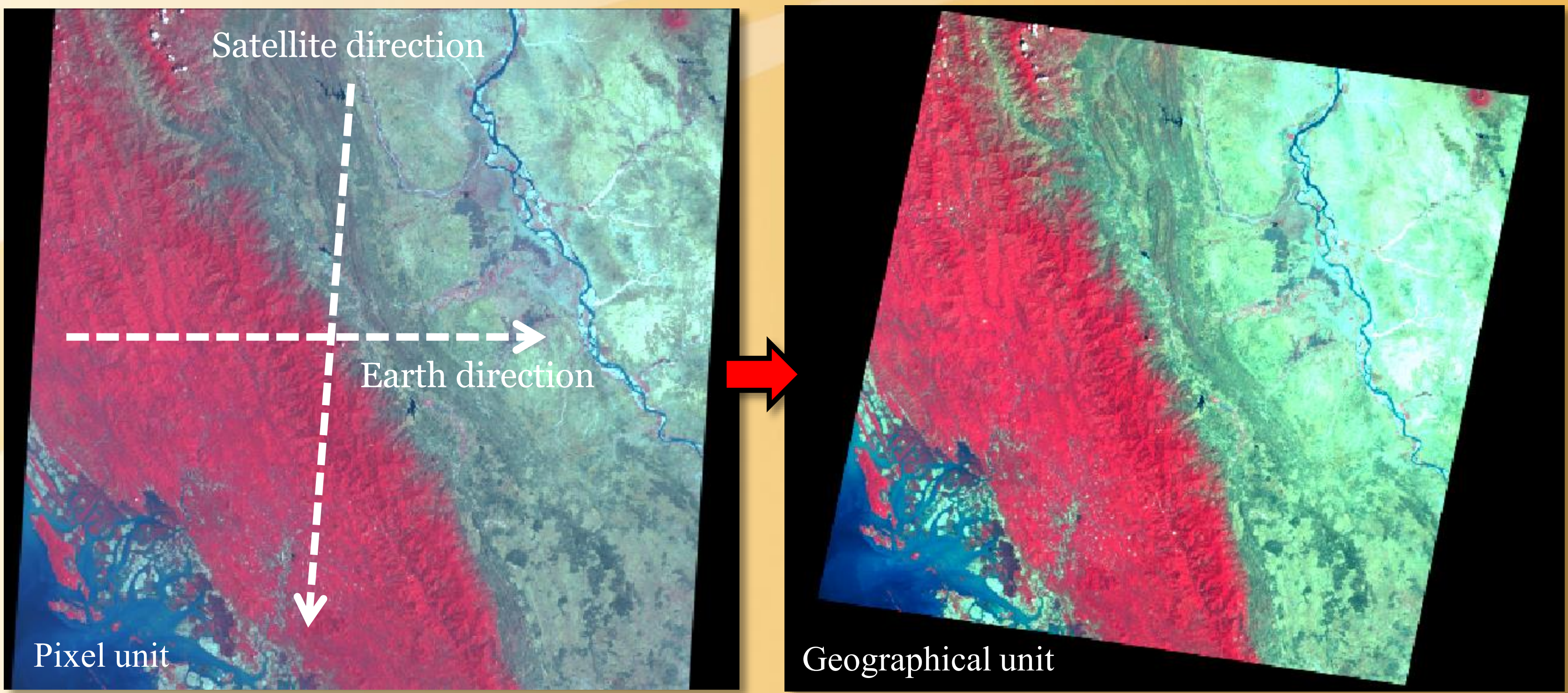


Small haze can be removed in Landsat TM/ETM.
But not clouds. Because B4 (IR) can penetrate the haze.

(b) Radiometric correction



(c) Geometric correction



Geometric distortion due to Earth rotation.

Methods of Geometric correction

1.Using satellite header file (satellite onboard GPS)

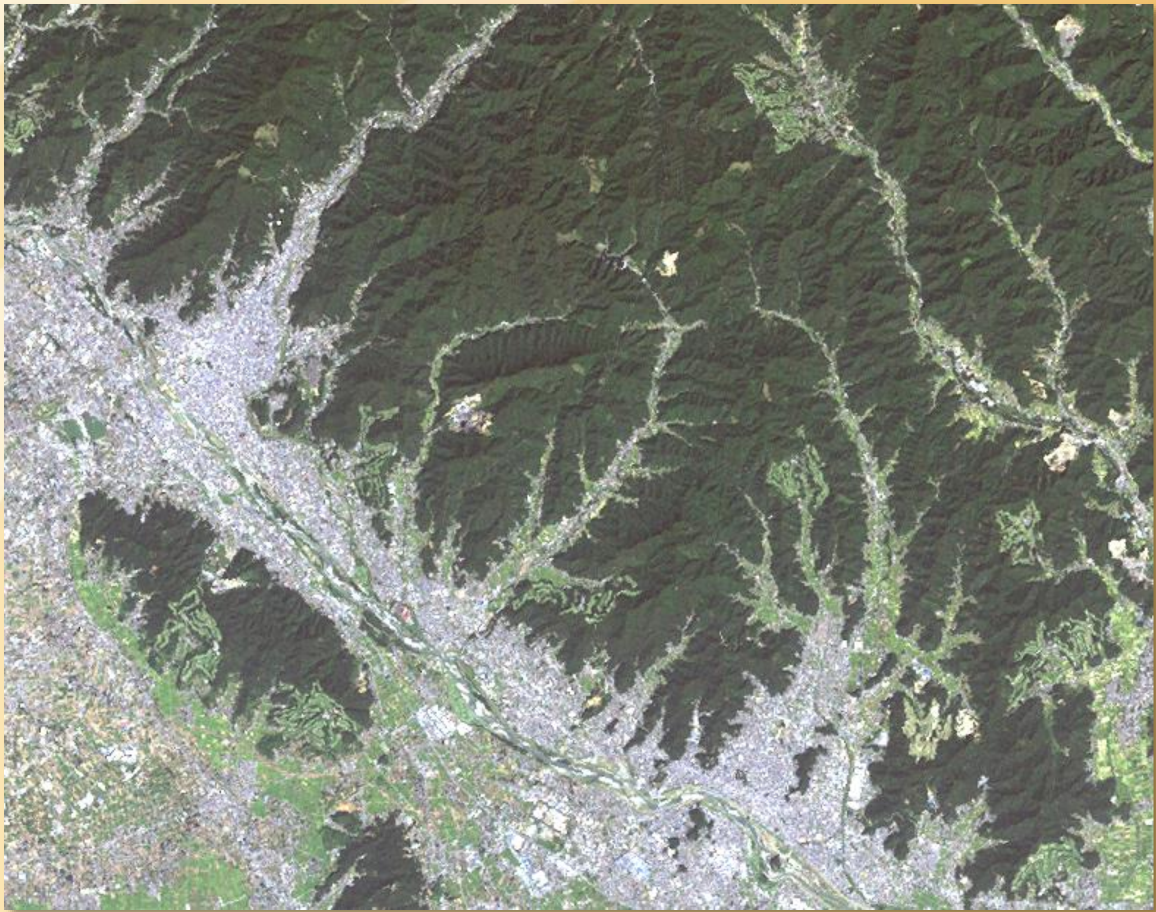
2.Image to image registration

3.Image to map registration

4.Manually entered GCPs (Ground Control Points)

3.2 Visual Interpretation (Band combination)

First step interpretation and to distinguish various land covers into different colors



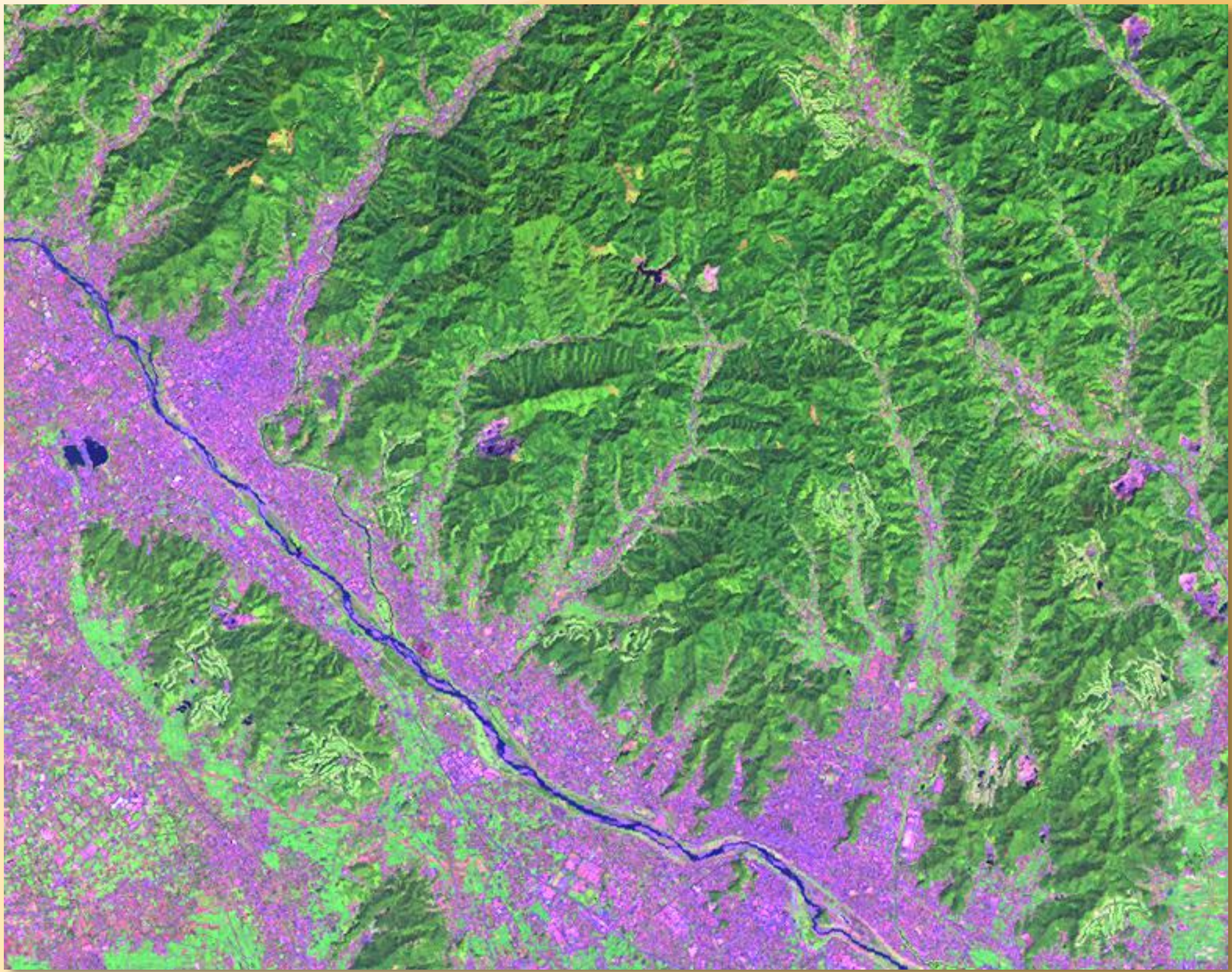
Landsat TM5 Tokyo (Ashikaga, Isezaki)

Example:

RGB 321 in Landsat TM/ETM gives natural color. Assign band 3 to red channel, band 2 to green channel and band 1 to blue channel in computer display.

To see landscape in realistic view.

3.2 Visual Interpretation (Band combination) *continued*



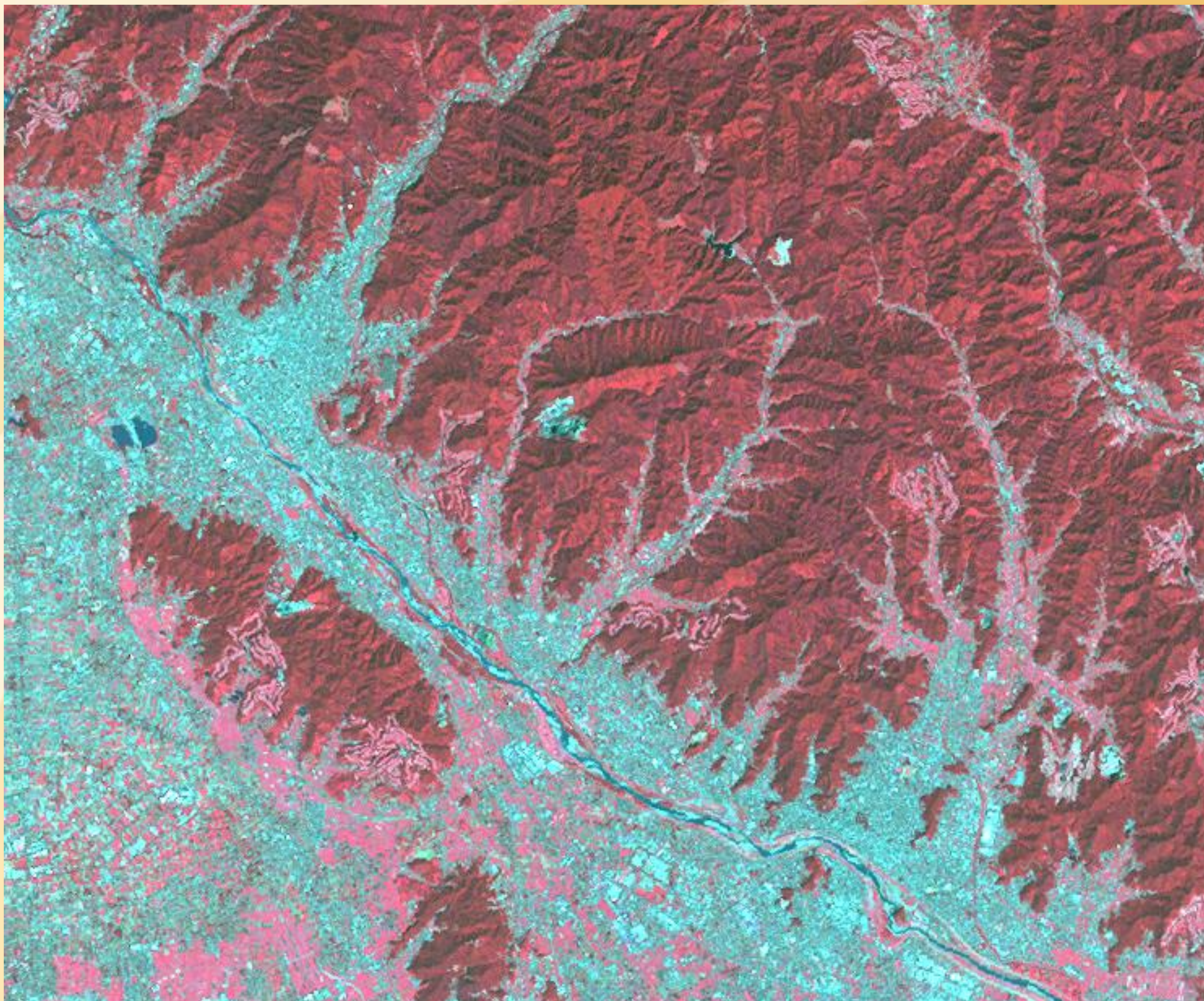
Landsat TM5 Tokyo (Ashikaga, Isezaki)

Example:

RGB 543 in Landsat TM/ETM gives false color. Assign band 5 to red channel, band 4 to green channel and band 3 to blue channel in computer display.

To discriminate between soil, vegetation and water.

3.2 Visual Interpretation (Band combination) *continued*



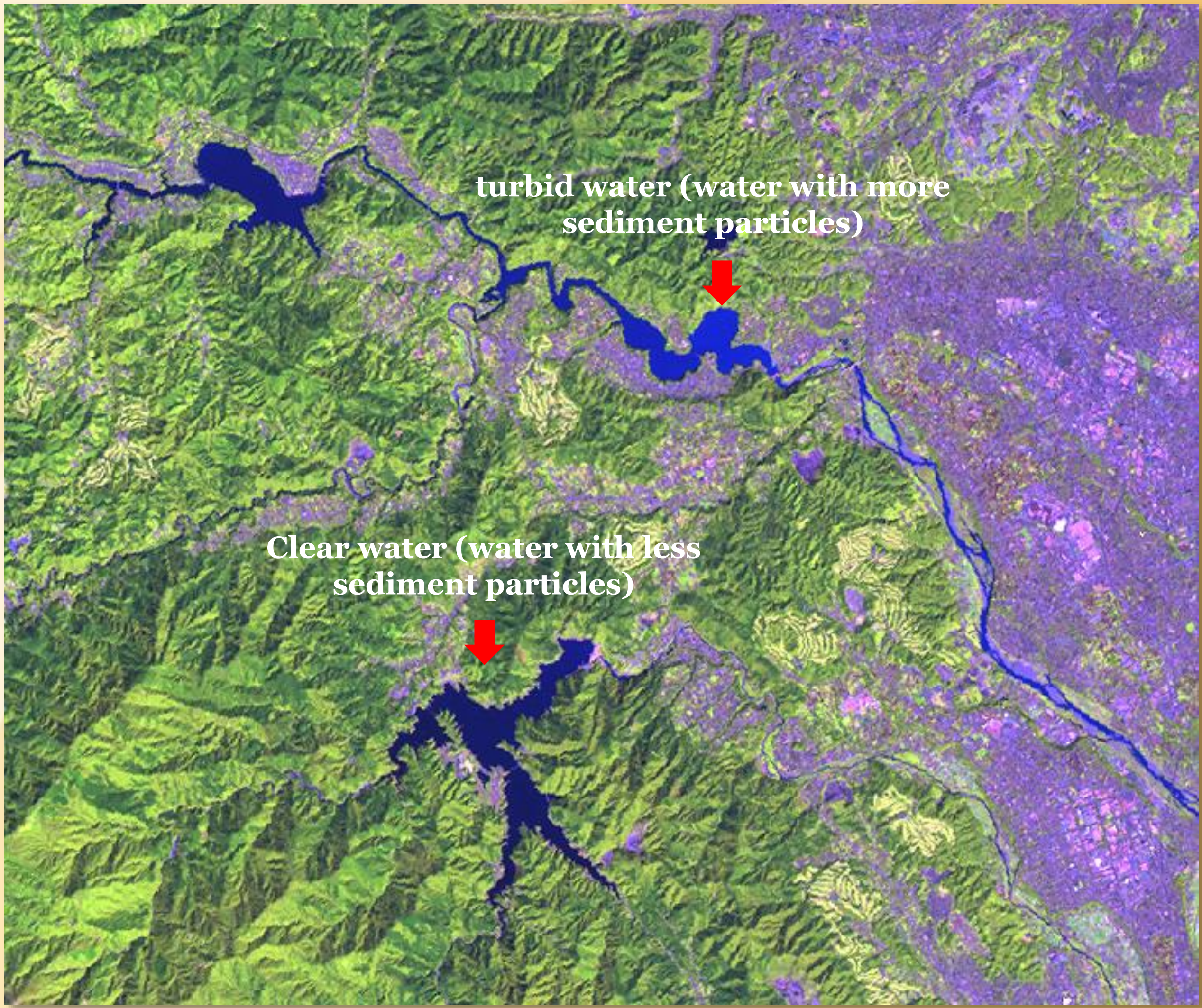
Landsat TM5 Tokyo (Ashikaga, Isezaki)

Example:

RGB 432 in Landsat TM/ETM gives false color. Assign band 4 to red channel, band 3 to green channel and band 2 to blue channel in computer display.

To determine vegetation stress and vigor.

3.2 Visual Interpretation (Band combination) *continued*



Landsat TM5 Tokyo (Hachioji)

Example:

RGB 541 in Landsat TM/ETM gives false color. Assign band 5 to red channel, band 4 to green channel and band 1 to blue channel in computer display.

To assess water quality. Turbid water gives bright blue and clear water gives dark blue.

3.3 Apply Algorithms

We can manipulate between bands (playing with DN Digital Numbers) and extract meaningful information.

(a) NDVI (Normalized Difference Vegetation Index)

Perhaps, well known and useful algorithm is NDVI (Normalized Difference Vegetation Index). Vegetation is low reflectance in Red band and high reflectance in Infrared band. By normalizing this two bands, we can measure vegetation stress and vigor.

General formula	NDVI = (Infrared – Red) / (Infrared + Red)
	The value is between +1 (vigor) ~ -1 (stress)
NOAA AVHRR	NDVI = (B2 – B1) / (B2 + B1)
Landsat TM/ETM	NDVI = (B4 – B3) / (B4 + B3)
IKONOS/QuickBird	NDVI = (B4 – B3) / (B4 + B3)

(b) NBR (Normalized Burn Ratio)

Landsat TM/ETM NBR = (B4 – B7) / (B4 + B7)

These two bands provide the best contrast between photosynthetically healthy and burned vegetation (Howard et al. 2002).

3.3 Apply Algorithms *(continued)*

Landsat TM5 Tokyo (Hanno)

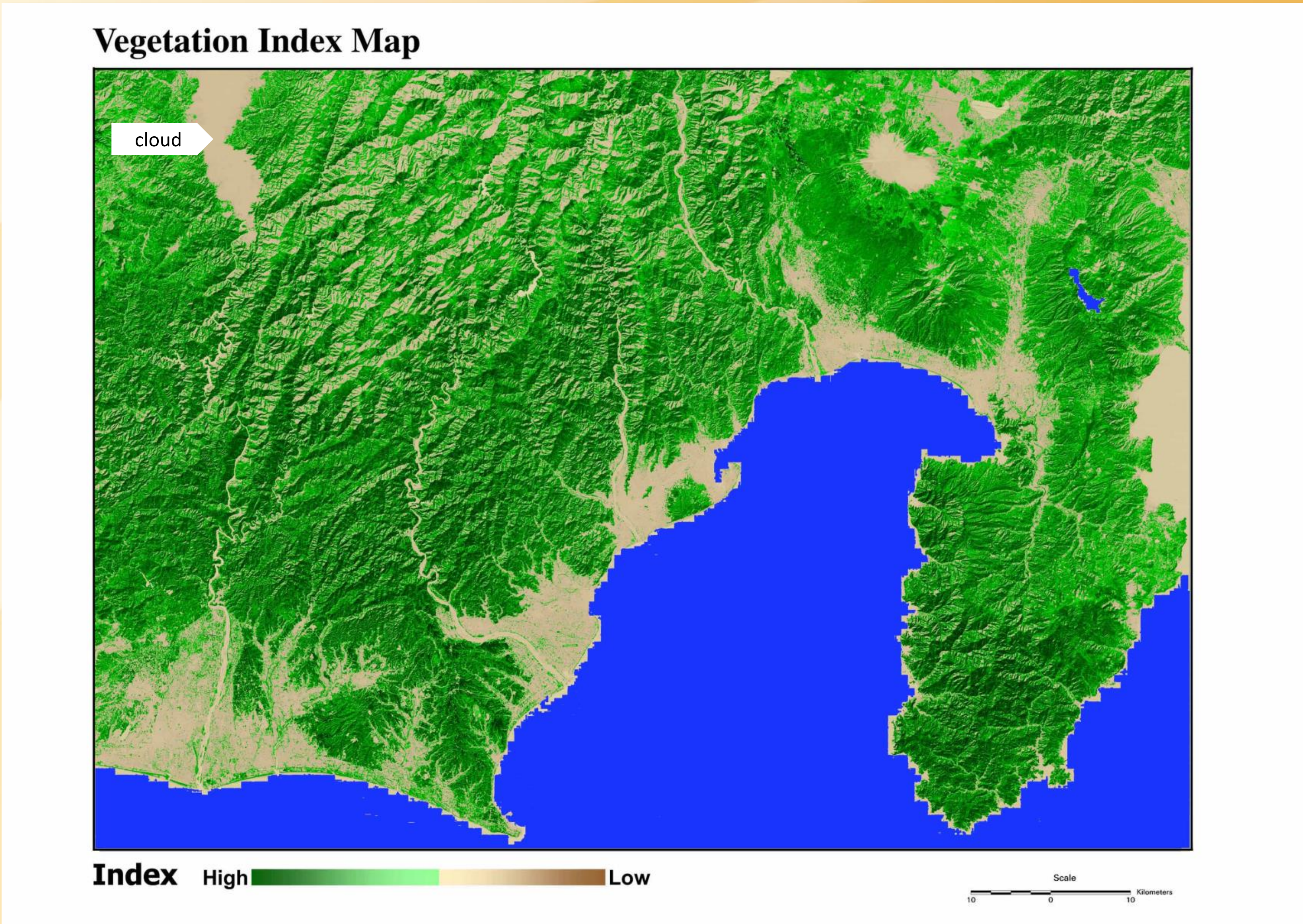
Example:

Vegetation index (NDVI) stretched to 8-bit.

-1 (low) +1 (high)

0 (low) 255 (high)

3.3 Apply Algorithms *(continued)*



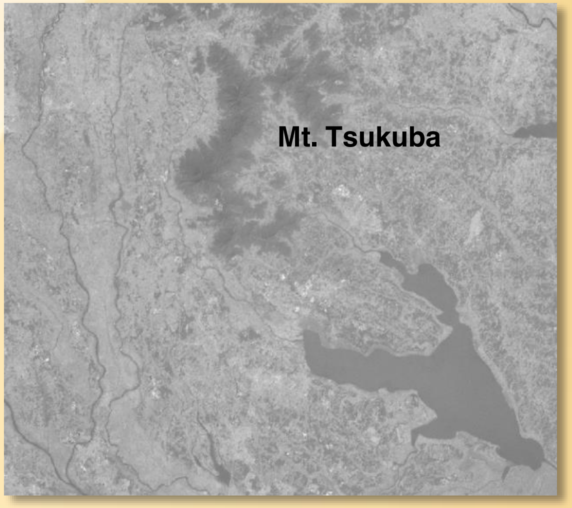
3.3 Apply Algorithms *(continued)*

(c) Surface Temperature

Some satellites carry thermal sensors. For example, Landsat TM/ETM, NOAA AVHRR ASTER, MODIS, etc.

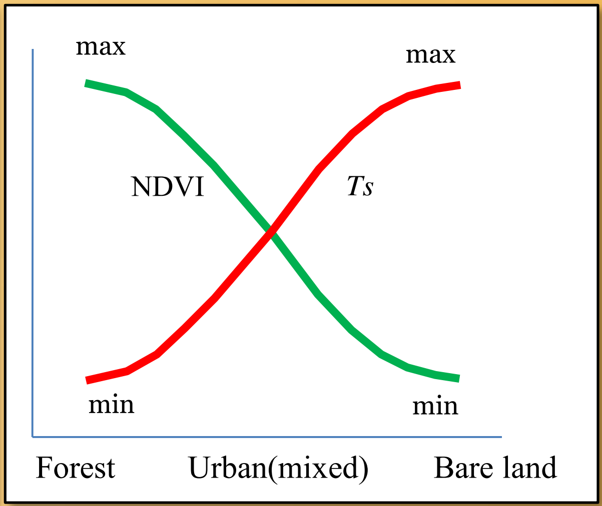
Thermal band records thermal emissive from the land surface objects. This band is good to study between surface temperature (*Ts*) and other land covers. For example, some researchers use surface temperature an NDVI to classify the land use land cover.

Thermal band spatial resolution is normally coarser than other bands because temperature does not change very well within the small area. Example: Landsat ETM thermal band spatial resolution is 60m compares to other bands (30m).



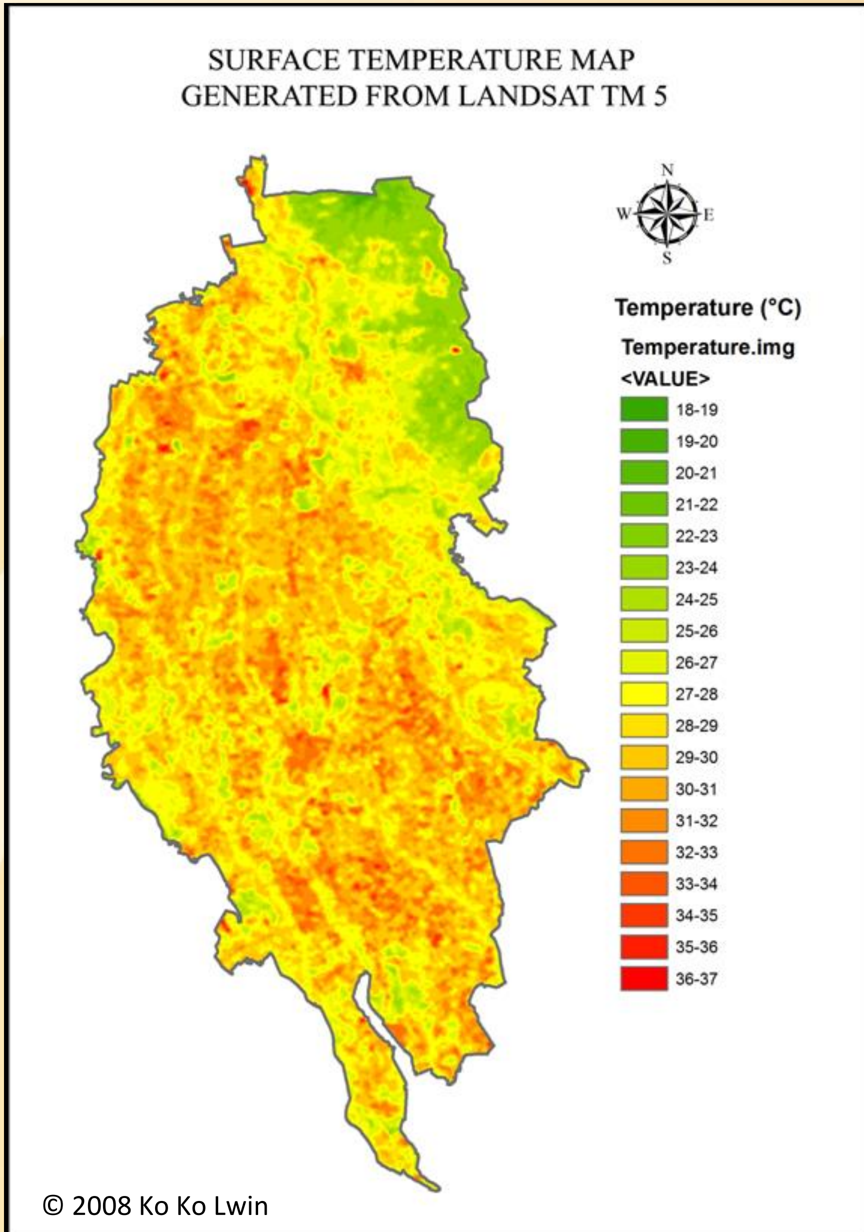
Landsat ETM Thermal Band 6 (Low gain)

Landsat ETM Thermal Band 6 (High gain)



NDVI vs. Ts

3.3 Apply Algorithms *(continued)*



Step1. Conversion of the Digital Number (DN) to Spectral Radiance (L)

L = LMIN + (LMAX - LMIN) * DN / 255

Where

L = Spectral radiance

LMIN = 1.238 (Spectral radiance of DN value 1)

LMAX = 15.600 (Spectral radiance of DN value 255)

DN = Digital Number

Step2. Conversion of Spectral Radiance to Temperature in Kelvin

$$T_B = \frac{K_2}{\ln\left(\frac{K_1}{L} + 1\right)}$$

Where

K_1= Calibration Constant 1 (607.76)

K_2 = Calibration Constant 2 (1260.56)

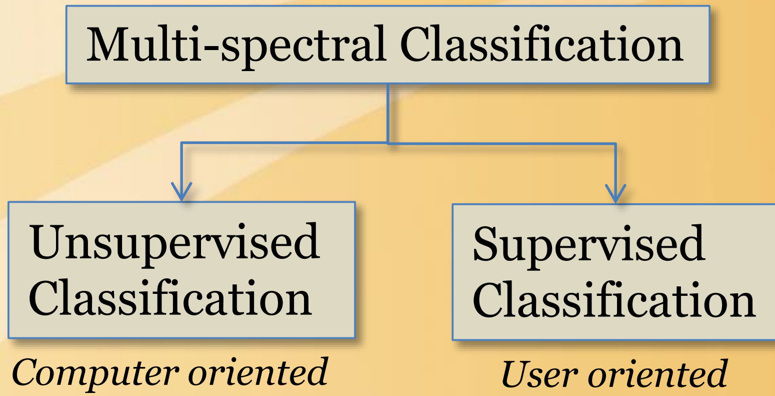
T_B = Surface Temperature

Step3. Conversion of Kelvin to Celsius

$T_B = T_B - 273$

Tsukuba City surface temperature map generated from Landsat TM5 satellite acquired by 1987-05-21, 11:00AM Local Time (JST)

3.4 Multi-spectral Classification



<u>Multi-spectral Classification</u>

The process of assigning individual pixels of an image to categories, generally on the basis of spectral reflectance characteristics. Two kinds of multi-spectral classifications.

<u>Unsupervised Classification</u>

Digital information extraction technique in which the computer assigns pixels to categories with no instructions from the operator. Also known as Isodata Classification.

<u>Supervised Classification</u>

Digital-information extraction technique in which the operator provides training-site information that the computer uses to assign pixels to categories.

3.4 Multi-spectral Classification *(continued)*

Unsupervised Classification (ERDAS Imagine Approach)

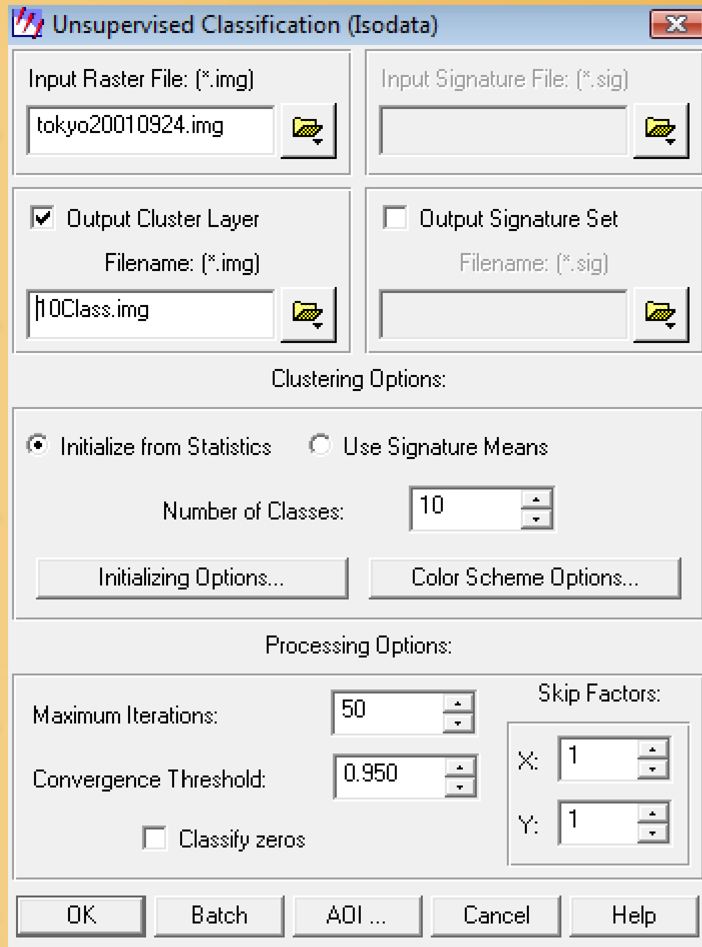
Insert classify image **1** →

Give out put file name **2** →

Set numbers of classes **3** →

Set Maximum iteration **4** →

Click OK to start to classify **5** →



Unsupervised classification is sometime use to know general clustered information from the image.

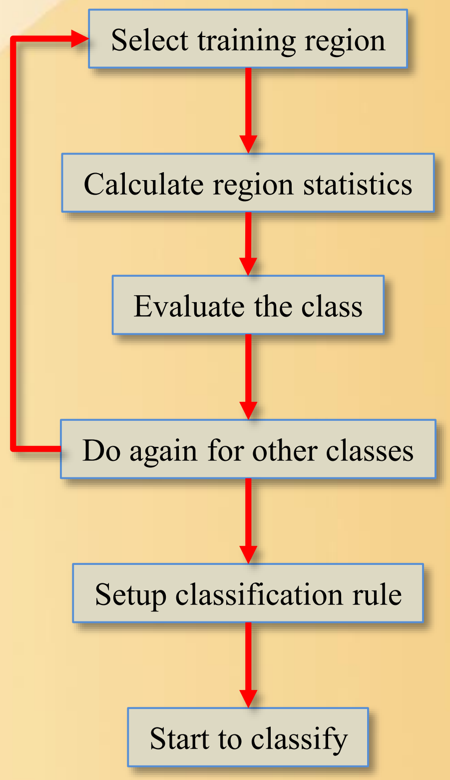
3.4 Multi-spectral classification *(continued)*

<u>Supervised Classification (ERDAS Imagine Approach)</u>

Signature:
Any characteristic or series of characteristics by which a material or object may be recognized in an image, photo, or data set.

Workflow

Signature collection steps



Draw polygon on an image

Compare signature statistics with other classes
-Avoid overlapping each other (signature separation)

Evaluate the selected class (pre-view)

Do again for other land use classes same as above

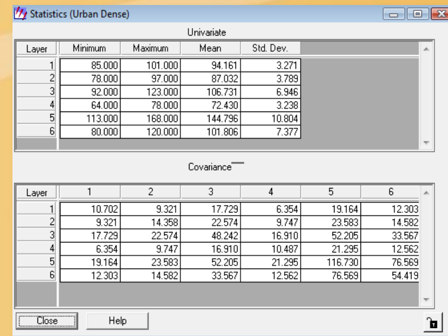
Setup classification rule
(e.g. Maximum Likelihood Classifier)

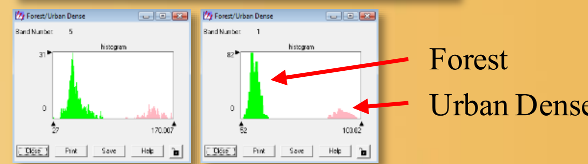
Start to classify an image

Step 1 : Select training region

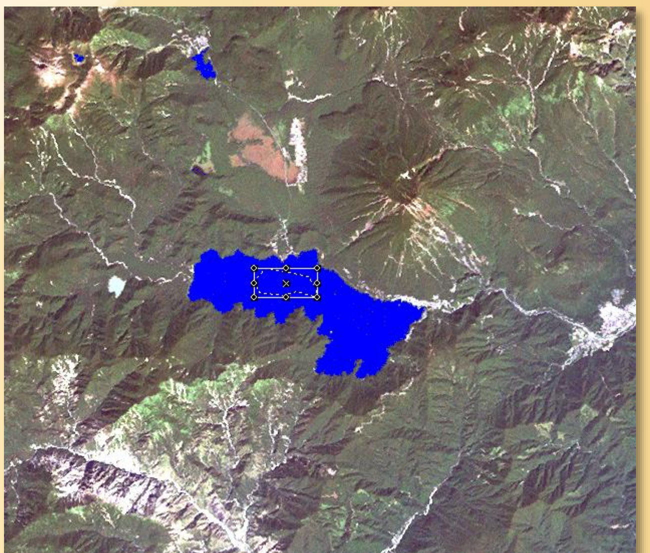
Training region or signature Signature Collection Tool

Step 2: Calculate region statistics

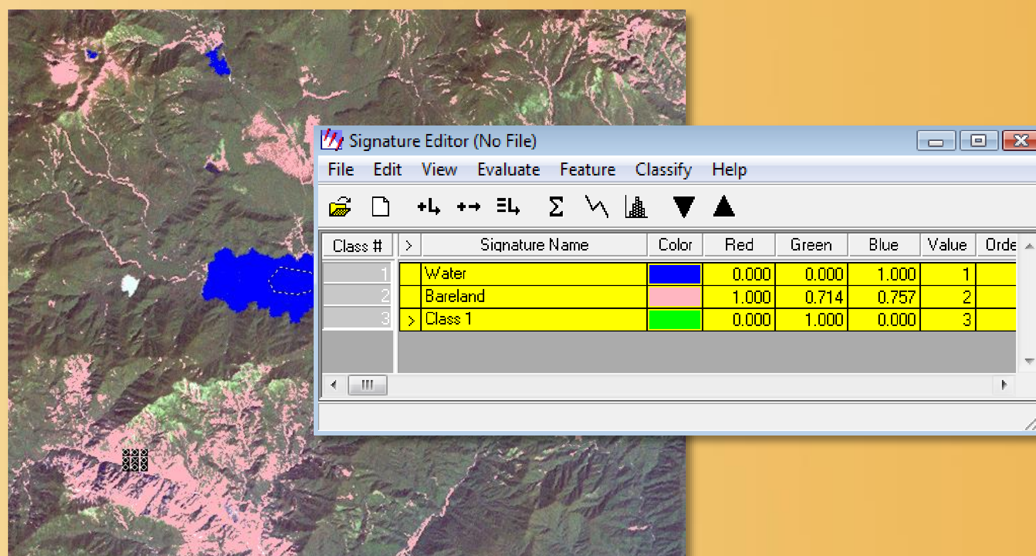




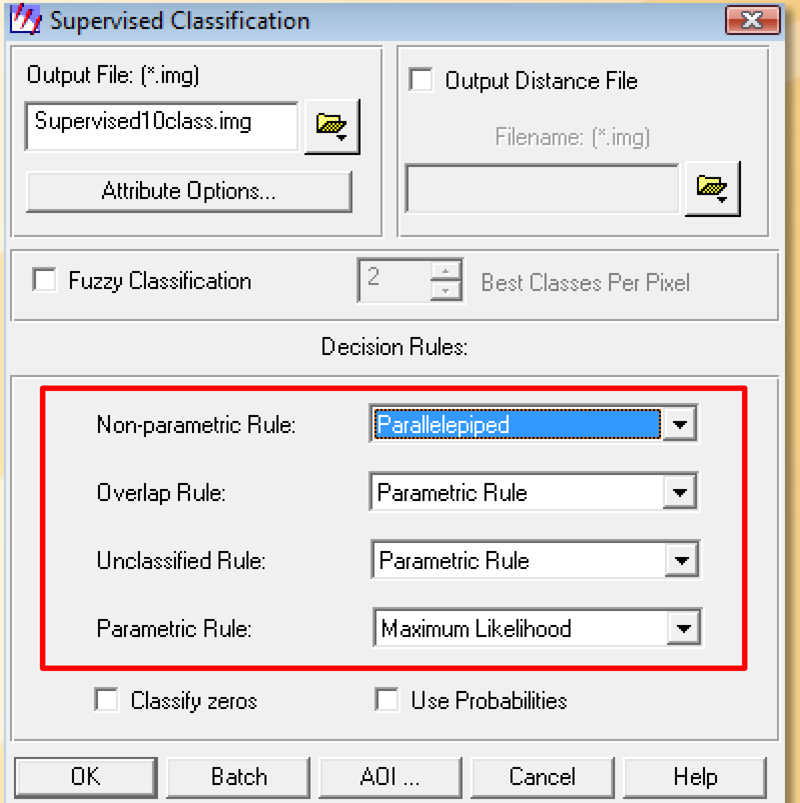
Step 3: Evaluate the class



Step 4: Do for other classes

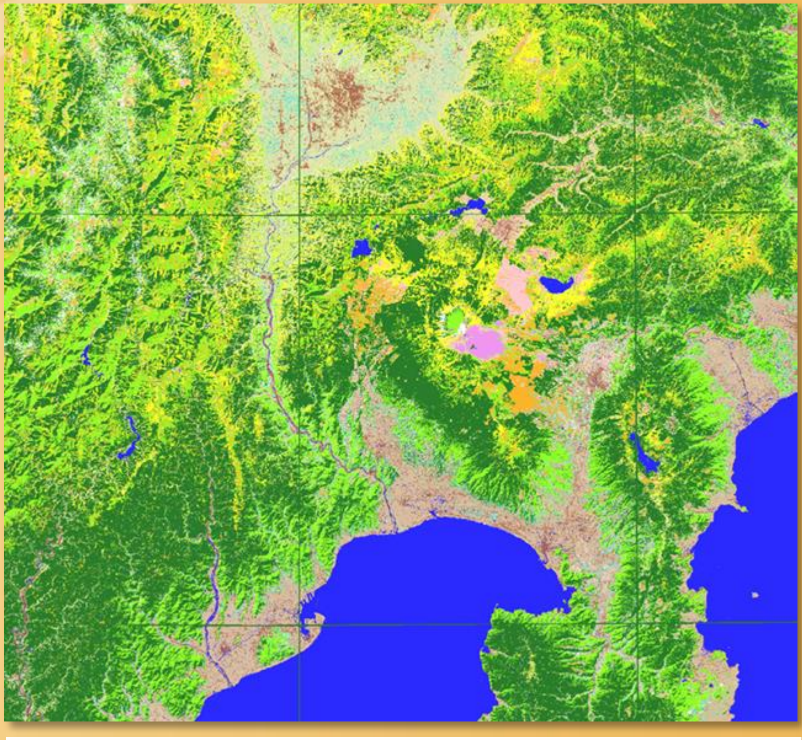


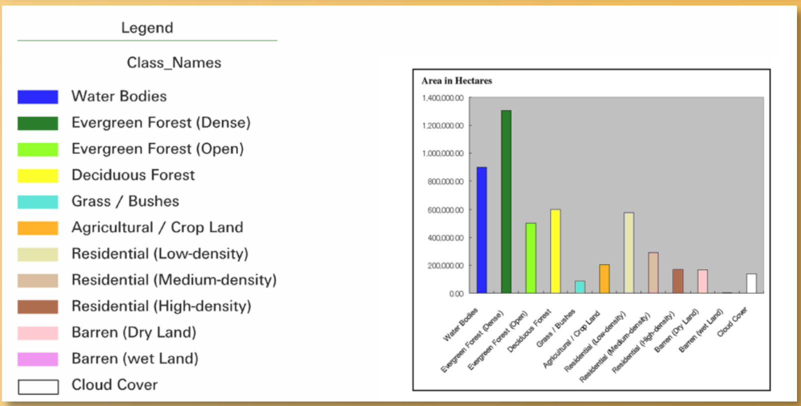
Step 5: Setup classification rule



Land use land cover is commonly used in spatial data modeling processes such as Hydrological Modeling, Soil Erosion and Land Degradation, Monitoring of Deforestation Process, Land Use Changes, etc.

Step 6: Start to classify

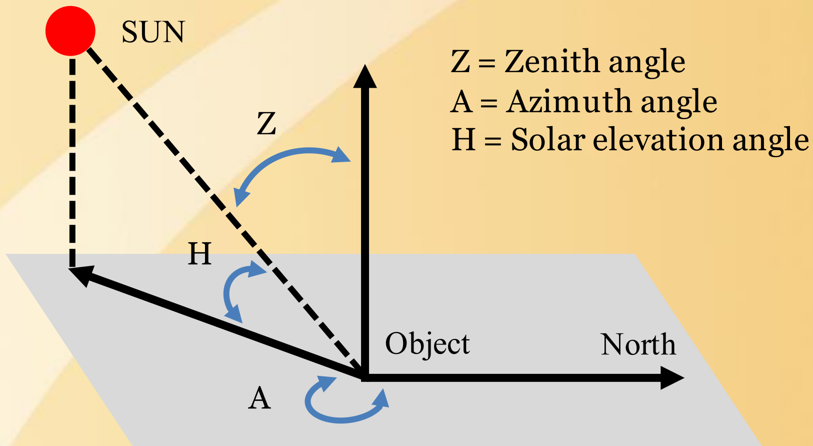


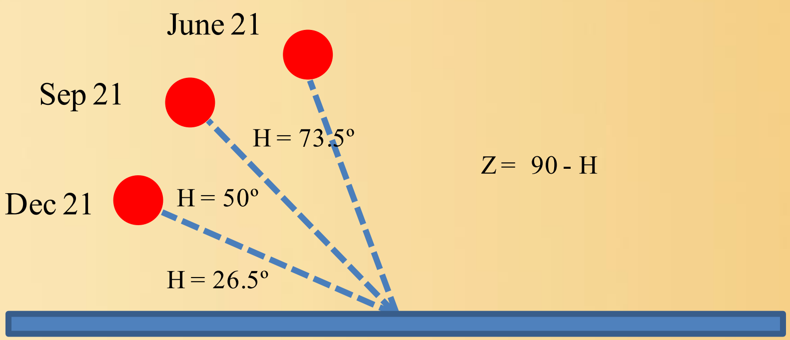


3.5 Scene Selection Criteria for Multi-spectral Classification

Scene should be:

1. Cloud free (if possible)
2. Plants growing season
3. Low solar zenith angle period





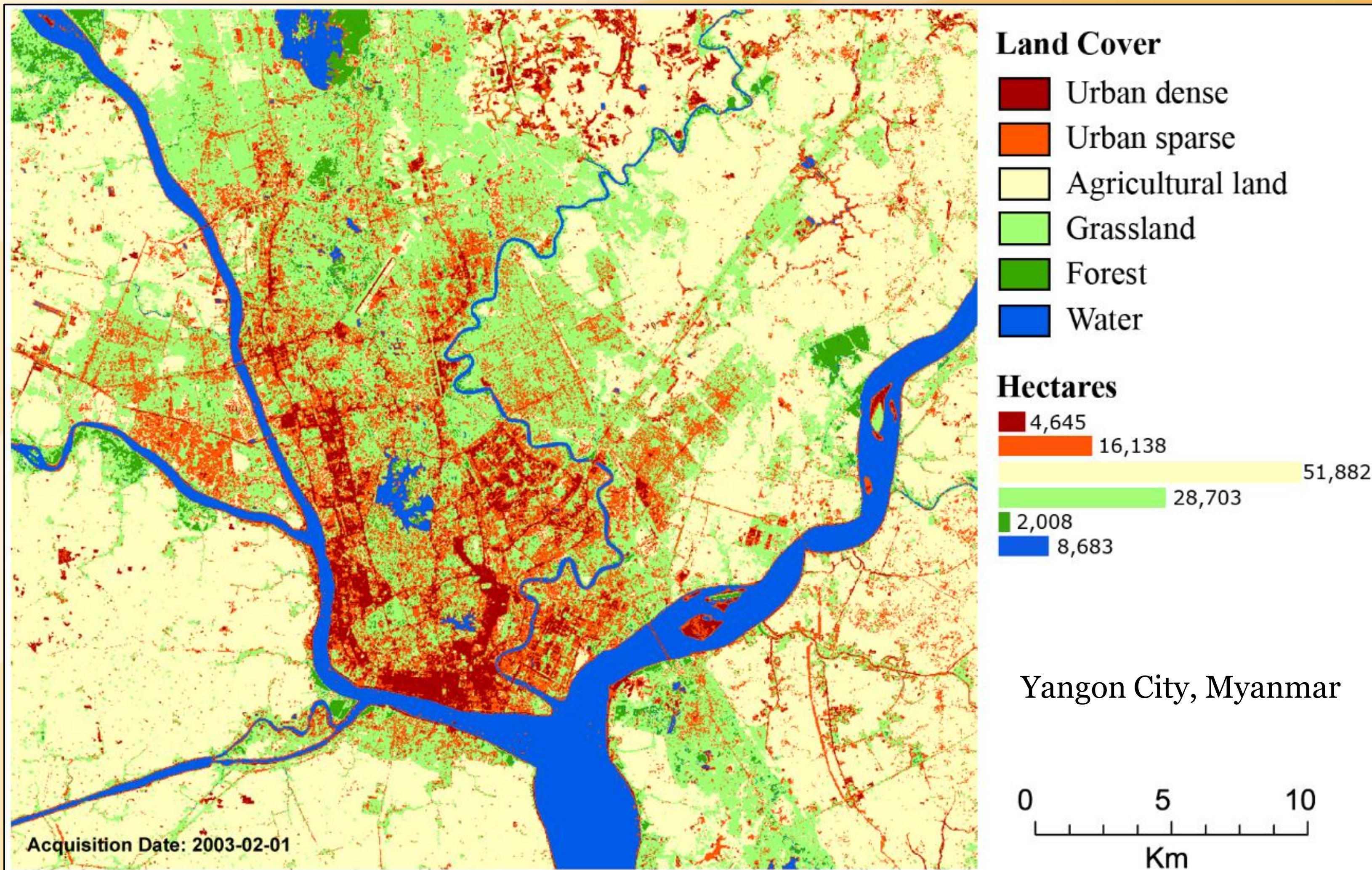
Normally effected in high rise mountain area and require additional "*Topographic Normalization*" process.

Image acquired in June

Image acquired in November

Part II: Remote Sensing Data Applications in GIS

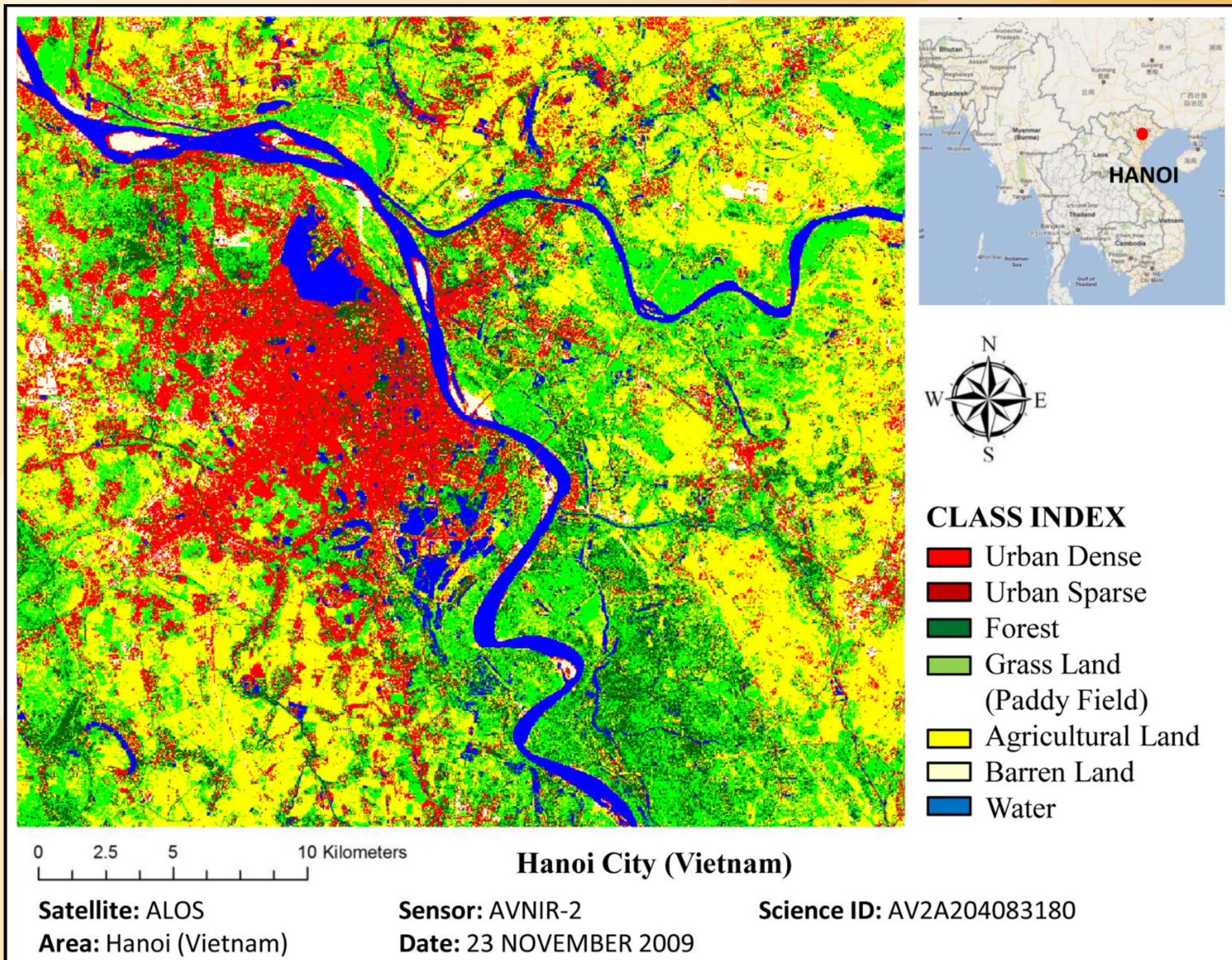
4.1 Land Cover Classification from Landsat ETM+



Land cover classification form Landsat ETM++

Source: Lwin, K. K. and Murayama, Y. (2013), Evaluation of land cover classification based on multispectral versus pansharpened Landsat ETM+ imagery, ***GIScience and Remote Sensing***, 50, 458-472.

4.2 Land Cover Classification from ALOS AVNIR-2



Land cover classification form ALOS AVNIR-2

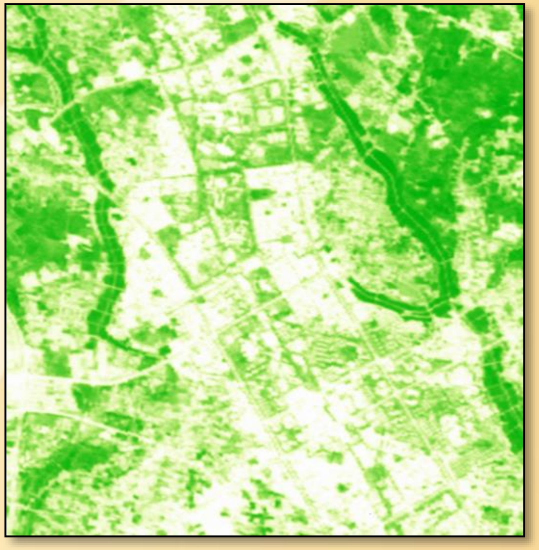
Source: Lwin, K. K. and Murayama, Y., (2011), Mapping the human settlement of South East Asia cities using ALOS AVNIR-2, ***Tsukuba Geoenvironmental Sciences***, 7: 13-17.

4.3 Urban Greenness (Eco-friendly Walk Score Calculator)

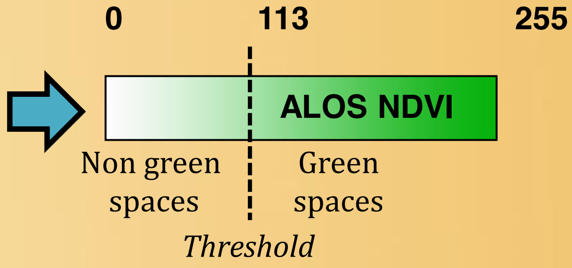
Web based interactive eco-friendly walk score calculator for Tsukuba City.

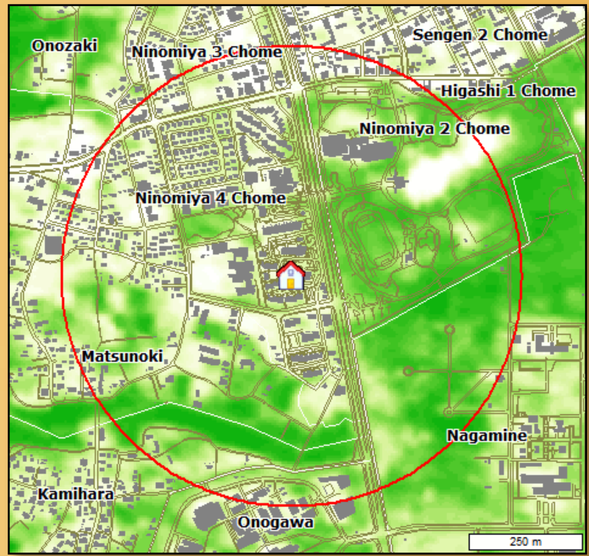
Eco-friendly Walk Score measures the degree of greenness (green density) by user defined geographic patterns based on Normalized Different Vegetation Index (NDVI).

NDVI was calculated from Advanced Land Observation Satellite ALOS AVNIR-2 sensor (ground resolution at 10m).



Green spaces:
Forest, paddy fields and grass lands
Non Green Spaces:
Bare lands, water surface, roads and building footprints



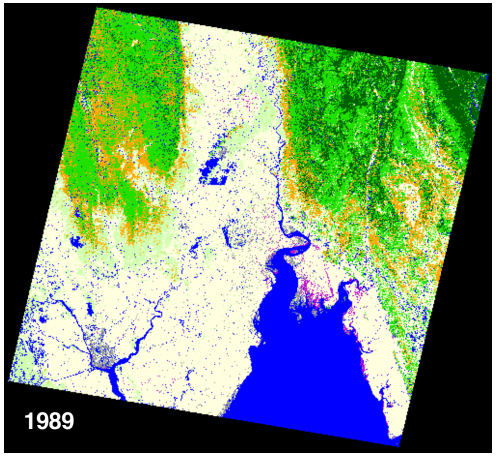


Search Radius: 500.06 m
Search Area: 78.56 Ha
Greenness Score: 55

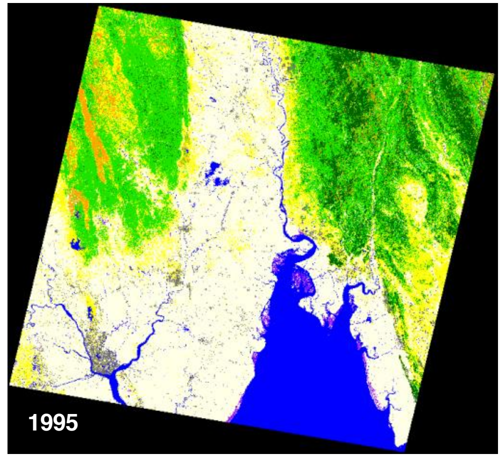
Choosing a Place to Live with GIS Project Homepage
Eco-friendly Walk Score Calculator for Tsukuba City
http://land.geo.tsukuba.ac.jp/ecowalk/default.aspx

Source: Lwin, K. K., & Murayama, Y., (2011), Modelling of Urban Green Space Walkability: Eco-friendly Walk Score Calculator, ***Computers, Environment and Urban Systems***, 35(5):408-420.

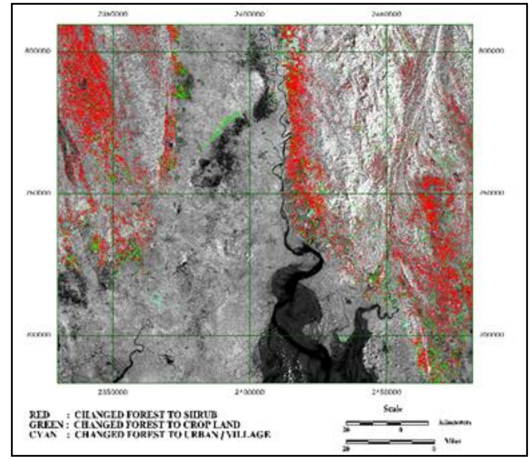
4.4 Monitoring of Deforestation Process



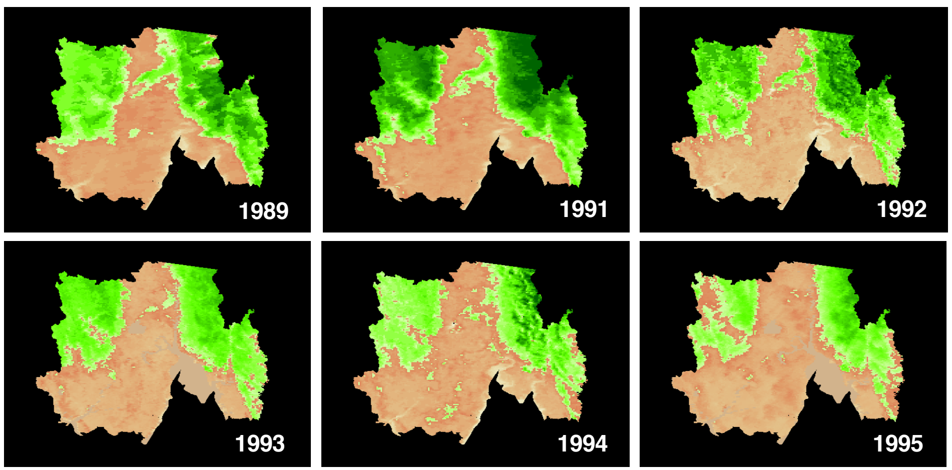
1988 Land use map

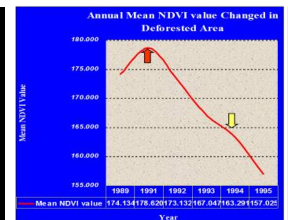


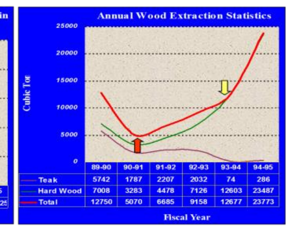
1995 Land use map



Deforested areas



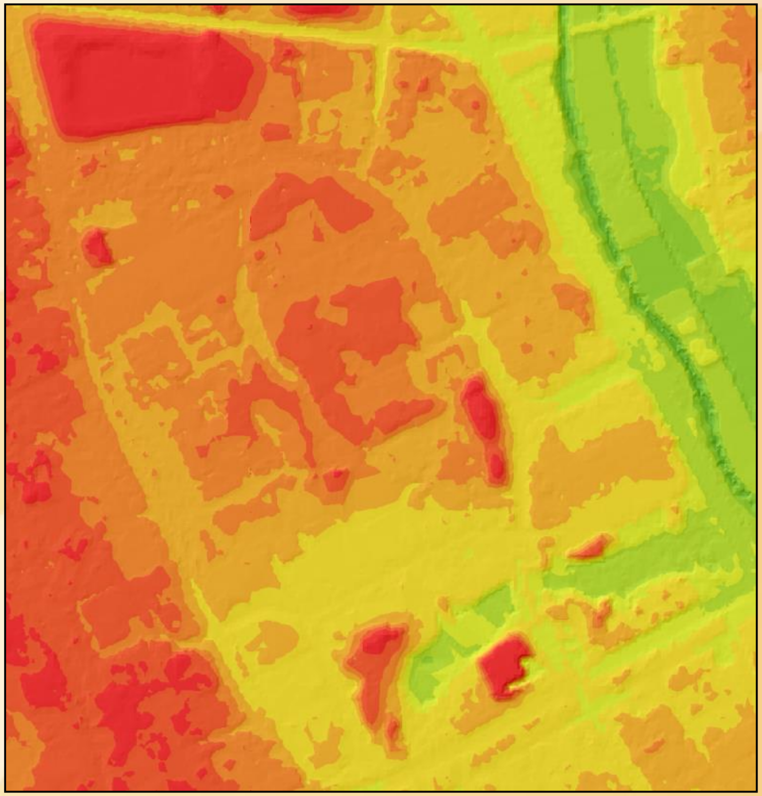




Monitoring of annual deforestation process using Landsat TM and NOAA AVHRR 10-day composite NDVI images (case study in Myanmar).

Source: Lwin, K. K. and Shibasaki, R., (1998), Monitoring and Analysis of Deforestation Process Using Remote Sensing and GIS: A Case Study of Myanmar, in: ***19th Asian Conference on Remote Sensing (ACRS)***, Manila, Philippines.

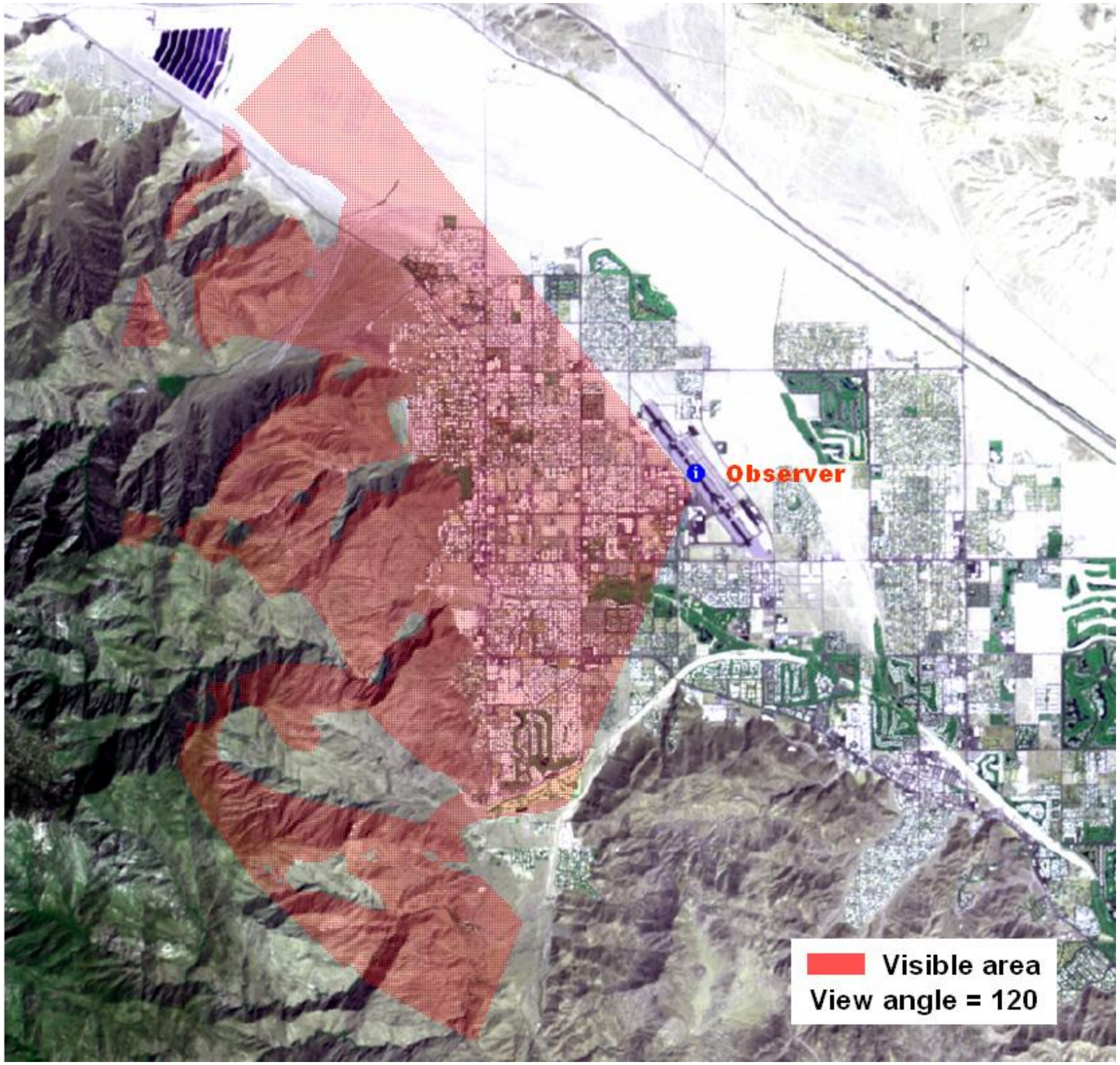
4.5 Surface Steepness Measurement from LIDAR

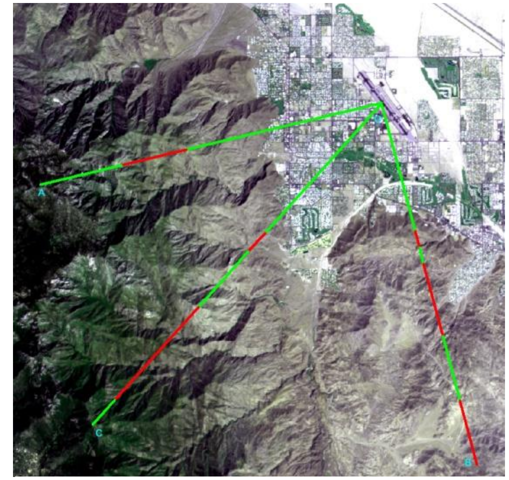


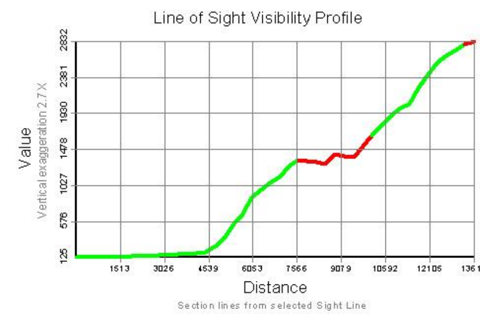
DTM and DSM generation form LIDAR data to measure terrain height and surface height (from sea level).
DTM = Digital Terrain Model
DSM = Digital Surface Model

Source: Lwin, K. K., Zhou, Y. and Murayama, Y., (2013), Identification of Bicycle Lanes Steepness from LIDAR Data, ***Tsukuba Geoenvironmental Sciences***, 8: 9-15.

4.6 Viewshed Analysis and Resort Site Selection

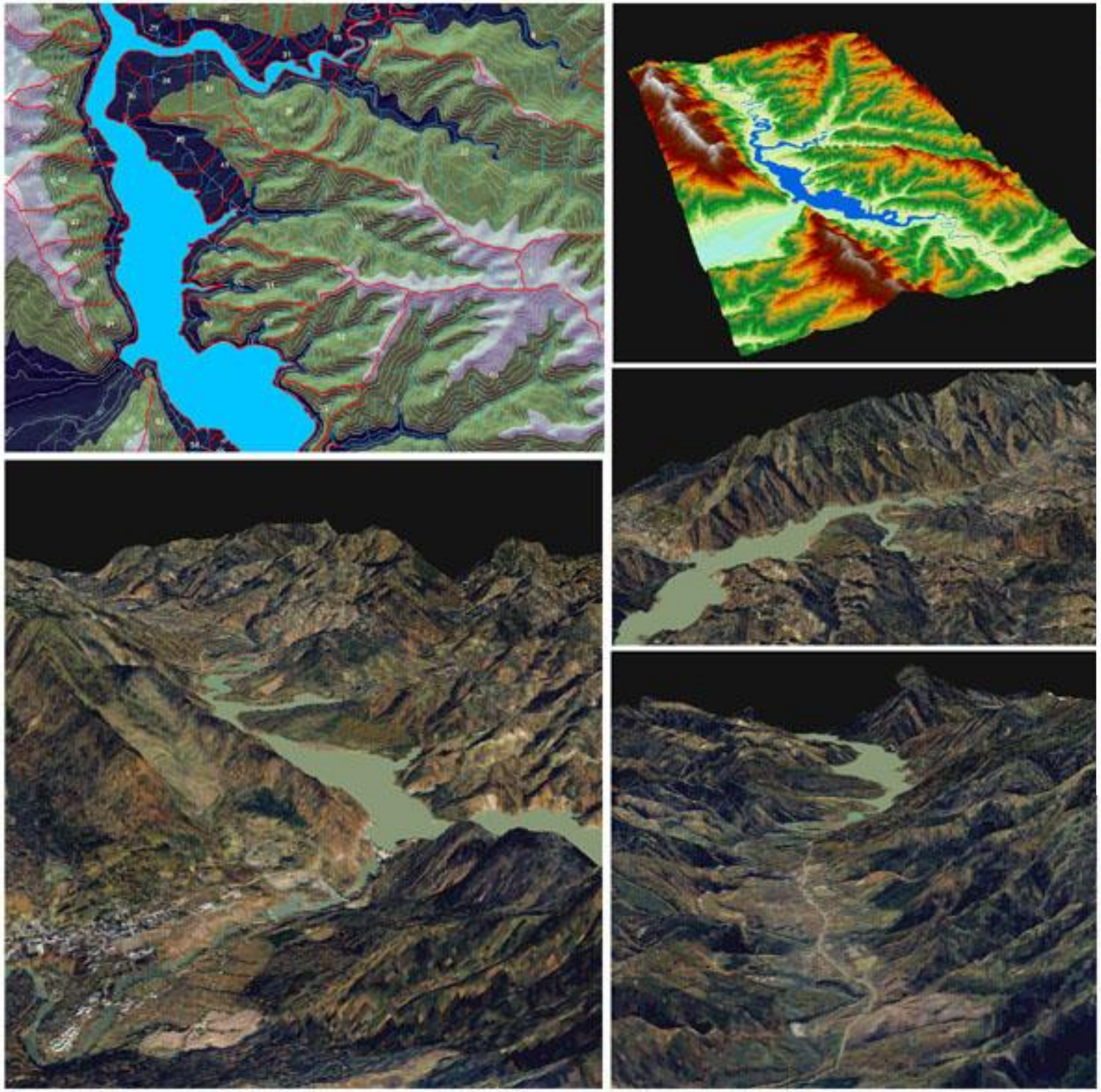






Viewshed and line of sight analysis based on DEM (Digital Elevation Model) and Aerial Photos for resort site selection. (Data source: ERDAS imagine)

4.7 Watershed and Environmental Impact Assessment



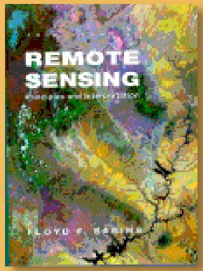
Watershed and environmental impact assessment. Aerial photos orthorectification, mosaicking and watershed delineated from DEM.

Appendix A Remote Sensing Learning Resources

BOOKS

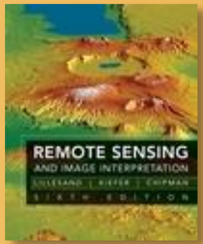
<u>Beginners</u>

Remote Sensing: Principles and Interpretations (Hardcover)
by Floyd F. Sabins (Author)



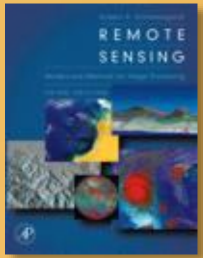
<u>Intermediate</u>

Remote Sensing and Image Interpretation (Hardcover)
by Thomas M. Lillesand (Author), Ralph W. Kiefer (Author), Jonathan W. Chipman (Author)



<u>Advanced</u>

The one who wants to develop own image processing algorithms…
Remote Sensing, Third Edition: Models and Methods for Image Processing (Hardcover) by Robert A. Schowengerdt (Author)



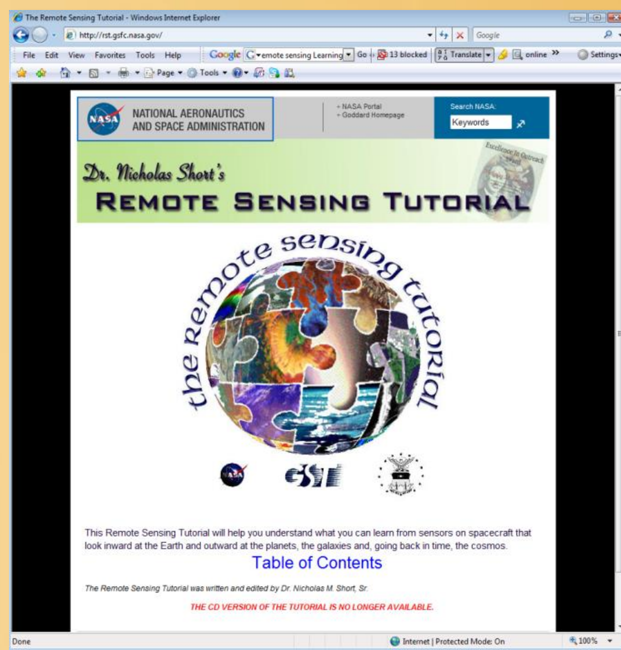
Appendix A Remote Sensing Learning Resources *(continued)*

Online Learning

CCRS Canada Centre for Remote Sensing
http://landmap.mimas.ac.uk/ipc/ccrs/fundam_e.html

NASA Remote Sensing Tutorial
http://rst.gsfc.nasa.gov/

TELSAT, Belgium
http://eoedu.belspo.be/en/guide/index.htm



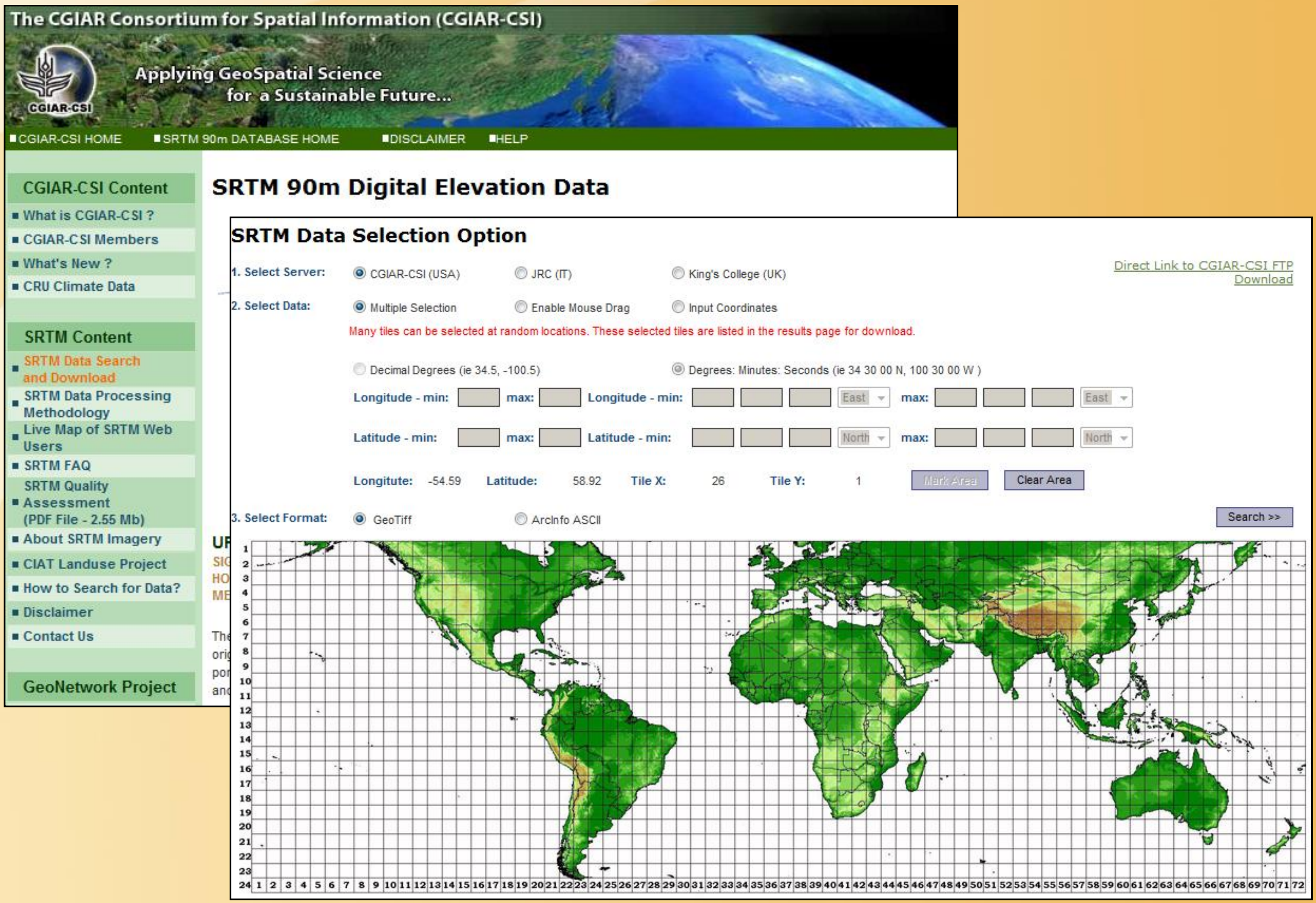
Appendix B Remote Sensing Data Resources

The GLCF is a center for land cover science with a focus on research using Remotely sensed satellite data and products to assess land cover change for local to global systems.
http://www.landcover.org/index.shtml ; http://www.landsat.org/ (Free)

Appendix B Remote Sensing Data Resources *(continued)*

SRTM 90m Digital Elevation Data (Free)

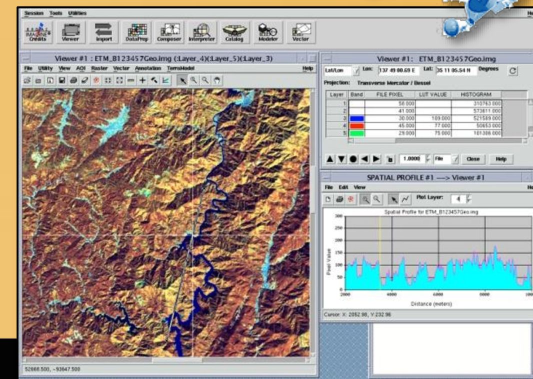
http://srtm.csi.cgiar.org/



Appendix C Remote Sensing Software Resources

<u>Freeware</u>

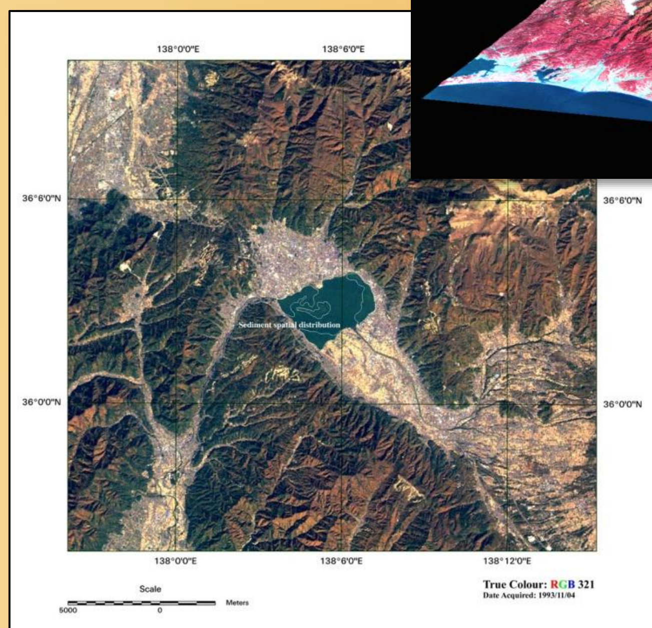
MultiSpec (A Multispectral Image Data Analysis System)
http://cobweb.ecn.purdue.edu/~biehl/MultiSpec/



<u>Commercial</u>

ERDAS Imagine
http://gi.leica-geosystems.com/LGISub1x33x0.aspx

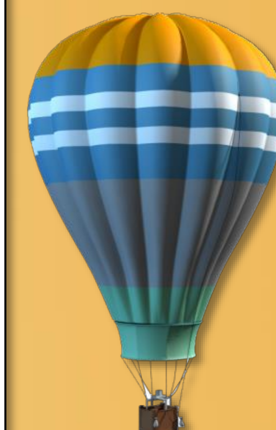
PCI Geomatics
http://www.pcigeomatics.com/



ENVI
http://rsinc.com/envi/

ER Mapper
http://www.ermapper.com/

IDRISI
http://www.clarklabs.org/



ISRO Indian Satellites List - Study Notes for SSC & Bank Exams in PDF

ISRO (Indian Space Research Oragnisation) has successfully launched GSAT – 11 into space today i.e., on 5th December 2018. This satellite is also called as "Big Bird" was launched from French space port Kourou in South America at 2.07 am (IST). The main purpose of GSAT-11 is to provide satellite based internet to remote places and will aid in providing internet connectivity in flights in India. It weights 5,854 kg is heaviest Indian Satellites that ISRO lauched into the orbit. This launch was the second attempt by ISRO after the first attempt which failed in May.

GSAT – 11 is the latest, next generation high throughout communication satellite that will accelerate the broadband service across the nation. Moreover, this satellite will be a great platform for new generation applications. It has cost about Rs.600 crore, and its lifespan is of 15 years.

Indian Satellite GSAT - 29

The Indian Space Research Organisation (ISRO) on 14th November 2018 also accomplished its fifth launch for the year by launching the **GSLV Mark III** rocket carrying the **GSAT-29** communication satellite from the Satish Dhawan Space Centre at Sriharikota. The communication satellite carries high throughput communication transponders in the Ka and Ku bands and its mission is to expand high-speed data transfer in the remote areas of India.

Before we look into the list of Indian Satellites, let's quickly learn few trivia about Satellites.

1. What is a Satellite?

Anything that orbits something else like the moon orbits the earth, is known as a satellite. They are used for diverse purposes such as weather forecasting, television broadcast, radio communications, internet communications, GPS, etc.

Generally, there are two types of satellites

- **Natural** (Moon orbiting the Earth)
- **Artificial** (International Space Station orbiting the Earth)

There are many natural and artificial satellites, out there in space, performing their functions and making our lives back on earth easy in some or the other way. The artificial satellites as the name suggests are man made satellites from different countries.

Let us look at the year - wise list of Indian Satellites with their purposes and also learn about the organisation responsible for the Indian Space programmes.

2. Types of Satellites

- **Geosynchronous**
  -It is placed in **geosynchronous** orbit.
  -It has an orbital period the same as the Earth's rotation period which means that it returns to the same position in the sky after each sidereal day.
- **Geostationary**
  -This is an earth-orbiting satellite.
  -It is placed at an altitude of approximately 35,800 kilometers (22,300 miles) directly over the equator.
  -It revolves in the same direction the earth rotates (west to east).

3. List of Indian Satellites

India has launched **106 satellites** of various types since its first in **1975.** The organisation responsible for Indian satellites is the **Indian Space Research Organisation (ISRO)**. Run through the list of Indian satellites, listed year-wise along with their purposes.

Indian Satellites	Launched Date	Purpose
Aryabhata	19th April 1975	- India's first satellite. - It was build to gain experience in building and operating a satellite in space.
Bhaskar	7th June 1979	- First experimental remote sensing satellite. - Carried TV and microwave cameras.
Rohini Technology Payload	10th August 1979	Intended for measuring in-flight performance of first experimental flight of SLV-3, the first Indian launch vehicle.
Rohini RS-1	18th July 1980	India's first indigenous satellite launch.

Rohini RS-D1	31st May 1981	Conducts remote sensing technology studies using a landmark sensor payload.
Ariane Passenger Payload Experiment	19th June 1981	First experimental communication satellite.
Bhaskara - II	20th November 1981	Second experimental remote sensing satellite.
INSAT-1A	10th April 1982	First operational multi purpose communication and meteorology satellite.
Rohini RS-D2	17th April 1983	Identical to RS-D1
INSAT-1B	30th August 1983	Earth observation satellite.
Stretched Rohini Satellite Series (SROSS-1)	24th March 1987	Carried payload for launch vehicle performance monitoring and for gamma ray astronomy.
IRS-1A	17th March 1988	First operational remote sensing satellite
Stretched Rohini Satellite Series (SROSS-2)	13th July 1988	Carried remote sensing payload of German space agency in addition to Gamma Ray Astronomy payload.
INSAT- 1C	21st July 1988	Same as INSAT-1A.
INSAT- 1D	12th June 1990	Identical to INSAT-1A.
IRS-1B	29th August 1991	- Earth observation satellite. - Improved version of IRS-1A
INSAT-2DT	26th February 1992	- It was a communications Satellite, earlier called as Arabsat. - After its retirement, it was placed in the Graveyard orbit
Stretched Rohini Satellite Series (SROSS-C)	20th May 1992	Carried gamma ray astronomy and astronomy payload.

INSAT- 2A	10th July 1992	First satellite in the second-generation Indian-built INSAT-2 series.
INSAT- 2B	23th July 1993	Second satellite in INSAT-2 series.
IRS-1E	20th September 1993	Earth observation satellite
Stretched Rohini Satellite Series (SROSS-C2)	4th May 1994	Identical to SROSS-C.
IRS-P2	15th October 1994	Earth observation satellite
INSAT-2C	7th December 1995	Has an additional capability such as mobile satellite service, business communication and television outreach beyond Indian boundaries.
IRS-1C	28th December 1995	Earth observation satellite
IRS-P3	21st March 1996	Carries remote sensing payload and an X-ray astronomy payload.
INSAT-2D	4th June 1997	Same as INSAT-2C.
IRS-1D	29th September 1997	Earth observation satellite.
INSAT-2E	3rd April 1999	Multipurpose communication and meteorological satellite.
Oceansat-1(IRS-P4)	26th May- 1999	- Carries an Ocean Color Monitor (OCM) and a Multi frequency Scanning Microwave Radiometer (MSMR).- Earth observation satellite.
INSAT-3B	21st March 2000	Multipurpose communication: business communication, developmental communication, and mobile communication.
GSAT-1	18th April 2001	Experimental satellite for the first developmental flight of Geosynchronous Satellite.

Technology Experiment Satellite (TES)	22nd October 2001	Experimental satellite to test technologies such as attitude and orbit control system, high-torque reaction wheels, new reaction control system, etc.
INSAT-3C	23rd January 2002	Designed to augment the existing INSAT capacity for communication and broadcasting and provide continuity of the services of INSAT-2C.
Kalpana-1 (METSAT)	12th September 2002	First meteorological satellite built by ISRO. Originally named METSAT. **Renamed after Kalpana Chawla**.
INSAT-3A	9th April 2003	Multipurpose satellite for communication, broadcasting, and meteorological services along with INSAT-2E and Kalpana-1.
GSAT-2	8th May 2003	Experimental satellite for the second developmental test flight of Geosynchronous Satellite.
INSAT-3E	27th September 2003	Communication satellite to augment the existing INSAT System.
RESOURCE SAT-1 (IRS-P6)	17th October 2003	- Earth observation/remote sensing satellite. - Intended to supplement and replace IRS-1C and IRS-1D.
EDUSAT	20th September 2004	India's first exclusive educational satellite.
HAMSAT	5th May 2005	Micro satellite for providing satellite-based amateur radio services to the national as well as the international community.
CARTOSAT-1	5th May 2005	Provides stereographic in-orbit images with a 2.5-meter resolution.
INSAT-4A	21st December 2005	Advanced satellite for direct-to-home television broadcasting services.
INSAT-4C	10th July 2006	Geosynchronous communications satellite.
CARTOSAT-2	10th January 2007	Advanced remote sensing satellite carrying a panchromatic camera capable of providing scene-specific spot images.
Space Capsule Recovery Experiment (SRE-1)	10th January 2007	Experimental satellite intended to demonstrate the technology of an orbiting platform for performing experiments in micro gravity conditions.

INSAT-4B	12th March 2007	Augments the INSAT capacity for direct-to-home (DTH) television services and other communications.
INSAT-4CR	2nd September 2007	It carried 12 high-power Ku-band transponders designed to provide direct-to-home (DTH) television services.
CARTOSAT-2A	28th April 2008	Earth observation/remote sensing satellite.
IMS-1 (Third World Satellite – TWsat)	28th April 2008	Low-cost micro satellite imaging mission.
Chandrayaan-1	22nd October 2008	Carries 11 scientific instruments built in India, USA, UK, Germany, Sweden and Bulgaria.
RISAT-2	20th April 2009	Radar imaging satellite used to monitor India's borders and as part of anti-infiltration and anti-terrorist operations.
ANUSAT	20th April 2009	- Carries an amateur radio and technology demonstration experiments. - Research micro satellite designed at Anna University.
Oceansat-2(IRS-P4)	23th September 2009	- Gathers data for oceanographic, coastal and atmospheric applications. - Continues mission of Oceansat-1.
GSAT-4	15th April 2010	Communications satellite technology demonstrator.
CARTOSAT-2B	12th July 2010	Earth observation/remote sensing satellite.
StudSat	12th July 2010	- First Indian pico-satellite (weighing less than 1 kg). - Developed by a team from seven engineering colleges from Karnataka and Andhra Pradesh.
GSAT-5P /INSAT-4D	25th December 2010	C-band communication satellite.
RESOURCESAT-2	20th April 2011	ISRO's eighteenth remote-sensing satellite
Youthsat	20th April 2011	Indo-Russian stellar and atmospheric satellite with the participation of university students.

GSAT-8 /INSAT-4G	21st May 2011	Communications satellite carries 24 Ku-band transponders and 2 channel GAGAN payloads operating in L1 and L5 band.
GSAT-12	15th July 2011	Extended C-band transponders to meet the country's growing demand for transponders in a short turn-around-time.
Megha-Tropiques	12th October 2011	Developed by India and France to track the weather.
Jugnu	12th October 2011	Nano-satellite weighing 3 kg developed by IIT Kanpur.
RISAT-1	26th April 2012	First indigenous all-weather Radar Imaging Satellite (RISAT-1), whose images will facilitate agriculture and disaster management.
SRMSAT	12th October 2011	Nano-satellite developed by SRM University.
GSAT-10	29th September 2012	India's advanced communication satellite is a high power satellite being inducted into the INSAT system.
SARAL	25th February 2013	The Satellite with ARGOS and ALTIKA (SARAL) is a joint Indo-French satellite mission for oceanographic studies.
IRNSS-1A	1st July 2013	It is one of the seven spacecraft constituting the IRNSS space segment.
INSAT-3D	25th July 2013	Meteorological Satellite with advanced weather monitoring payloads.
GSAT-7	30th August 2013	Advanced multi-band communication satellite dedicated for military use.
Mars Orbiter Mission (MOM)	5th November 2013	Also known as **Mangalyaan** is India's first Mars orbiter.
GSAT-14	5th January 2014	Twenty-third geostationary communication satellite of India to augment the In-orbit capacity of Extended C and Ku-band transponders.

IRNSS-1B	4th April 2014	Second satellite of the IRNSS.
IRNSS-1C	15th October 2014	Third satellite of the IRNSS.
GSAT-16	7th December 2014	Twenty-fourth communication satellite of India configured to carry a total of 48 communication transponders.
IRNSS-1D	28th March 2015	Fourth satellite of the IRNSS
GSAT-6	27th August 2015	Communication satellite.
Astrosat	28th September 2015	India's first dedicated multi wavelength space Observatory.
GSAT-15	10th November 2015	Communications satellite carries communication transponders in Ku-band and a GPS Aided GEO Augmented Navigation (GAGAN) payload operating in L1 and L5 bands.
IRNSS-1E	20th January 2016	Fifth satellite of the IRNSS.
IRNSS-1F	10th March 2016	Sixth satellite of the IRNSS.
IRNSS-1G	28th April 2016	Seventh and final satellite of the IRNSS.
Cartosat-2C	22nd June 2016	Earth observation/remote sensing satellite.
SCATSAT-1	26th September 2016	Miniature satellite to provide weather forecasting, cyclone prediction, and tracking services to India.
RESOURCESAT-2A	15th February 2017	Remote Sensing satellite intended for resource monitoring.
CARTOSAT-2D	15th February 2017	Highest number of satellites launched by a single launch vehicle (104 satellites)
PSLV-C38 / Cartosat-2 Series Satellite	23rd June 2017	The Cartosat will provide remote sensing services for about five years.

PSLV-C40/Cartosat-2 Series Satellite Mission	12th June 2018	Providing high resolution scene specific spot imageries.
PSLV-C41/IRNSS-1I	12th April 2018	Navigation satellite constellation
NovaSAR, S1-4	16th September 2018	Intended for forest mapping, land use & ice cover monitoring, flood & disaster monitoring.
GSAT-29	14th November 2018	Aims at providing high-speed bandwidth to Village Resource Centres (VRC) in rural areas.
HysIS(Hyper spectral Imaging Satellite)	29th November 2018	Aims at studying the earth's surface in the visible, near infrared and shortwave infrared regions of the electromagnetic spectrum.

4. Indian Satellites - Important Facts

- ISRO was formed on the Independence Day, 1969 by Dr. Vikram Sarabhai.
- SLV-3 was India's first indigenous satellite launch vehicle. The director of this project was APJ Abdul Kalam.
- India is the only country to have reached Mars in the first attempt.
- The satellite Aryabhata got its name by Indira Gandhi and was launched by the Soviet Union.
- India has set a national record by successfully launching a rocket carrying 20 satellites, including 13 from the US, last year in June.

Try to remember list of Indian Satellites was helpful to you. For more such informative articles visit the links provided below!

General Awareness	
Oscar Winners 2018	National Film Awards
Indian Navy Day	Top 10 Most Populated Cities in the world

As we all know, practice is the key to success. Therefore, boost your preparation by starting your practice now.

Solve Practice Questions for Free

Furthermore, chat with your fellow aspirants and our experts to get your doubts cleared on Testbook Discuss:

Go to Testbook Discuss!

perties of Stars

In the previous module we learnt how the light elements were formed soon after the Big Bang. However our bodies are not composed purely of hydrogen and helium. The other elements that we see within us and around us must have been manufactured since that time. It turns out that the factories for the heavier elements are the stars.

Before we begin the story of the creation of heavy elements in stars we need to understand a few basic observable properties of stars.

Imagine a SETI astronomer sat at the controls of a radio telescope poised to search for radio transmissions from distant stars. One of the first decisions to be made is where the telescope should be pointed; this process of constructing a target list is one of the most important tasks in any programme of observational astronomy. Table 1 lists characteristics of the first two stars ever investigated in a SETI programme (Frank Drake's Project Ozma in 1960). In fact these two stars are still on the target list of the SETI Institute's current major programme Project Pheonix, their popularity stemming from the fact that they are the two nearest Sun-like stars.

Table 1. Characteristics of two popular target stars for SETI observations.

Name	RA	Dec	V magnitude	Spectral type	Distance
Tau Ceti	1h 44m 4.1s	-15d 56m 15s	3.5	G8V	11.8 l.y.
Epsilon Eridani	3h 32m 55.9s	-9d 27m 30s	3.7	K2V	10.7 l.y.

The table lists several characteristics which we will learn about during this course. In this module we will address the definition of magnitudes and spectral types.

e magnitude system for measuring brightnesses

The column headed **V magnitude** in Table 1 refers to the apparent brightness of the stars. The magnitude system for comparing stellar brightnesses was introduced by Hipparchus in the first century BC when he classified stars into six 'magnitude classes', first magnitude stars being the brightest, sixth magnitude ones the faintest.

This was formalised by Pogson in 1854 and is now defined by a system which has the following properties:

1. A difference of 5 magnitudes is exactly a factor 100 in brightness. This was Pogson's approximation to the difference in brightness between Hipparchus' magnitude 1 and magnitude 6 stars.
2. Somewhat contrary to expectation, a smaller or more negative magnitude means the star is brighter. For example: a star of magnitude 18 is much fainter than a star of magnitude 6; Sirius has a magnitude of -1.4 and is the brightest star in the sky other than the Sun; you should see from Table 1 that Tau Ceti is slightly brighter than Epsilon Eridani.
3. The scale is logarithmic i.e. a difference of 5 magnitudes equates to a multiplicative change of 100 in brightness. Hence a difference of 1 magnitude is a multiplicative change in brightness of about 2.512 (because 2.512 multiplied by itself 5 times equals 100 i.e. $2.512^5=100$ or equivalently $100^{1/5}=2.512$.

The relationship between the energy output of a star and its apparent magnitude

Consider a star at a distance *d* which has a total power output of *L* (i.e. the energy it radiates per second summed over all wavelengths); *L* is known as the **bolometric luminosity** of the star and is measured in Watts (Joules per second). The corresponding bolometric **flux** *F* of radiation at the Earth measured in W m^{-2} is then just given by

$$F = \frac{L}{4\pi d^2} \quad (1)$$

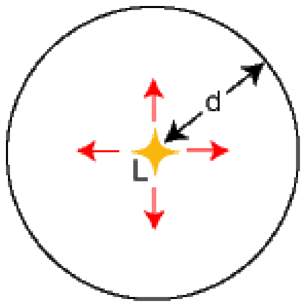


Figure 1. Star radiating luminosity *L* into a sphere of radius *d*.

because this energy is spread out over the surface of a sphere of radius *d* which has surface area $4\pi d^2$, Figure 1.

The luminosity is independent of distance and is therefore telling us something about the star's intrinsic brightness whilst the flux is providing information on the star's apparent brightness as seen from the Earth. So, if we could move a star farther away from us, its intrinsic brightness (luminosity) would not change but its apparent brightness (flux) would be reduced. Equation 1 quantifies the famous inverse square law.

The formal definition of the apparent magnitude *m* of a star in terms of its flux *F* is

$$m = -2.5 \log_{10} F + C \tag{2}$$

where *C* is a constant. The factor -2.5 and the value of *C* are chosen to approximate the scale to that of Hipparchus and to define the zero point (i.e. the flux which corresponds to zero magnitude).

It follows from the definition above that for two stars (subscripts 1 and 2)

$$m_1 - m_2 = -2.5 \log_{10} \left(\frac{F_1}{F_2} \right) \; . \tag{3}$$

In other words the difference in their magnitudes relates directly to a ratio of apparent brightnesses i.e. fluxes.

The faintest stars which one can see with the naked eye under optimum observing conditions have visual magnitudes of about 6 (this is known as the naked eye limiting magnitude). With large telescopes and modern detectors it is possible to reach magnitudes as faint as 28.

We can substitute for *F* in equation (2) to find

$$m = -2.5 \log_{10} \left(\frac{L}{4\pi d^2} \right) + C = -2.5 \log_{10} L + 5 \log_{10} d + D \tag{4}$$

where *D* is another constant. Note that this magnitude measures the apparent brightness of a star and is therefore termed the **apparent magnitude** as it is a function of the star's distance *d*. A lowercase *m* is always used to represent an apparent magnitude.

The absolute magnitude and the distance modulus

It is also useful to define an **absolute magnitude** which depends only on the star's intrinsic brightness. Then a comparison of two stars' absolute magnitudes would allow you to immediately say which of the two were actually the brighter and it wouldn't matter how far away each were. The absolute magnitude *M* (note the use of a capital M) is defined to be the apparent magnitude a star would have if it were moved to a standard distance *d* = 10 parsecs. Hence, from equation (4), absolute magnitude is given by

$$M = -2.5 \log_{10} L + 5 \log_{10} 10 + D = -2.5 \log_{10} L + 5 + D \ . \qquad (5)$$

The difference between the apparent and absolute magnitudes of a star is then just given by subtracting equation (5) from equation (4):

$$m - M = 5 \log_{10} d - 5 \ . \qquad (6)$$

This difference depends only on a star's distance *d* (measured in parsecs) and is usually referred to as **the distance modulus**.

Table 2 lists characteristics of some typical astronomical objects.

Table 2. Characteristics of several astronomical objects.

Star	Apparent magnitude	Distance (parsecs)	Absolute magnitude	Luminosity (relative to Sun)
Sun	-26.8		4.8	
Full Moon	-12.6			
Venus	-4.4			
Sirius	-1.44	2.64	1.45	22.5
Arcturus	-0.05	11.25	-0.31	114
Vega	0.03	7.76	0.58	50.1
Spica	0.98	80.40	-3.55	2250
Barnard's Star	9.54	1.82	13.24	1/2310

Proxima Centauri	11.01	1.30	15.45	1/17700

Is Sirius brighter or fainter than Spica (a) as observed from Earth and (b) intrinsically? What is the distance modulus of Sirius? Why should this be a negative number?

Click for answer

It is worth noting that the Sun is actually fairly bright compared to a typical star. Figure 2 shows that a typical star in the solar neighbourhood has an absolute magnitude of between 10 and 12. The Sun itself at an absolute magnitude of 4.8 lies in the column labelled 6 (magnitudes in the range 4-6).

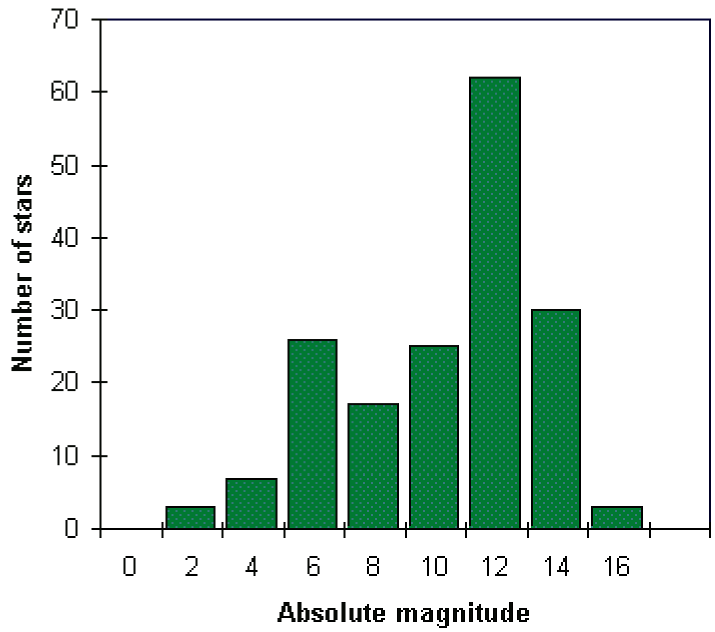


Figure 2. The frequency distribution of the absolute magnitudes of all stars within 10 parsecs of the Sun (from the Hipparcos database).

Magnitude systems and filters

In our discussion above, apart from a brief reference to the concept of bolometric luminosity, we didn't consider whether the flux we were measuring from a star was the light collected over a wide range of wavelengths or just from a narrow range e.g. a specific colour. In fact, astronomers have to be clear about what wavelength range a magnitude refers to.

If we could gather the light from a star across the whole range of the electromagnetic spectrum i.e. radio to gamma rays, then we would be measuring a **bolometric** magnitude. Of course, it is virtually impossible to measure a bolometric magnitude, partly because of absorption of some wavelengths in the Earth's atmosphere, but also because detectors/telescopes are not equally sensitive across all wavelengths. Therefore any given observation only detects some fraction of the total flux arriving at the Earth. In practise, the observer uses a filter to isolate some waveband of interest, this then determines the magnitude system.

The discussion is here confined to the area around the optical part of the electromagnetic spectrum. So-called narrow-band filters have typical widths of about 50 Angstroms (an Angstroms is 10^{-10}m, so this is equivalent to 5 nm), whilst broad-band filters are about 1000 Angstroms wide. The commonest system of broadband filters, due to Johnson and Morgan, is the system of UBV filters, see Figure 3.

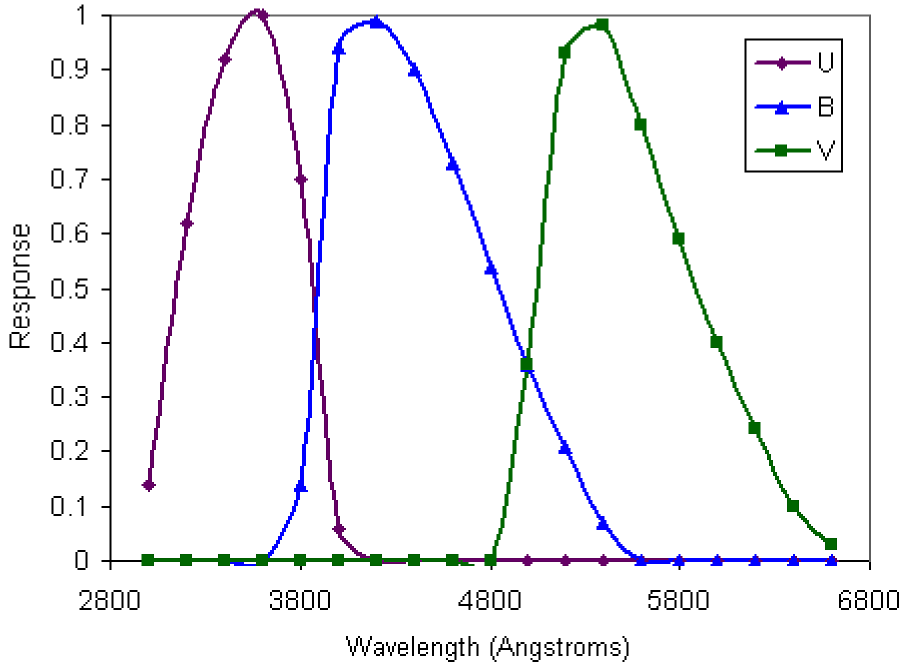


Figure 3. The UBV filter system. The three curves show the relative response as a function of wavelength for the U, B and V filters. For example, we see that the V filter only transmits light of wavelengths between about 4800 and 6800 Angstroms and is most sensitive to light of about 5400 Angstroms.

A V-band magnitude is then referred to as m_V or just V, for an apparent magnitude, and M_V for an absolute magnitude. V stands for visual, B for blue and U for ultraviolet. Note that the equations in the previous section which were written in terms of unspecified magnitudes could have been written down in exactly the same form for V-magnitudes or any other bandpass as long as one remembers the luminosity (in the case of absolute magnitudes) or flux (in the case of apparent magnitudes) is that measured in the same bandpass.

The UBV system was originally defined by Johnson & Morgan with a specific telescope/filter/instrument combination. Now observations are calibrated with respect to a set of standard stars whose magnitudes in this system are defined. The wavelengths of peak transmission of various filters in the extended system are: U-360 nm, B-440 nm, V-550 nm, R-700 nm, I-900 nm, J-1.25 μm, K-2.2 μm, L-3.4 μm, M-5.0 μm, N-10.2 μm and Q-21 μm.

Of course, a star observed through a filter which transmits only some fraction of the light incident upon it will appear fainter than if all its light were being measured. This difference is quantified as the **bolometric correction** (BC), which is defined as the difference between the bolometric magnitude and the V magnitude via

$$BC = M_{bol} - M_V \ ; \tag{8}$$

it is therefore always negative.

The constant C in equation (2) determining the zeropoint of the magnitude scale was originally defined so that the star Vega had visual and blue apparent magnitudes $m_B = m_V = 0$ but has since been redefined so that the Sun has visual and blue absolute magnitudes $M_B = 5.48$ and $M_V = 4.83$.

e colours of stars

The colour of a star can be parameterised by taking the ratio of brightnesses at two different wavelengths. In other words, when we say a star is particularly red we are actually comparing how bright it is in the red part of the spectrum to how bright it is in the blue part. Since the magnitude scale is logarithmic this ratio of brightnesses is equivalent to taking the difference of two magnitudes, see equation (3). Astronomers call a difference of two magnitudes from different wavebands a colour index. For example,

$$\text{B-V Colour Index} = m_B - m_V = B - V \ . \tag{9}$$

Another commonly used colour index is $U - B$.

Note that cooler (and therefore redder) stars have larger values of $B - V$. For example, Figure 4 depicts continuous spectra of two stars, one much hotter than the other. You should recognise these as blackbody spectra emitted because a star is a hot, largely opaque, glowing ball of gas. The central wavelengths at which the blue (B) and visual (V) magnitudes are measured are indicated. We see that the hot star is brighter in the B region than in the V region of the spectrum; therefore, its B magnitude is smaller than its V magnitude, and it has a negative $B - V$ color index. The opposite is true for the cool star.

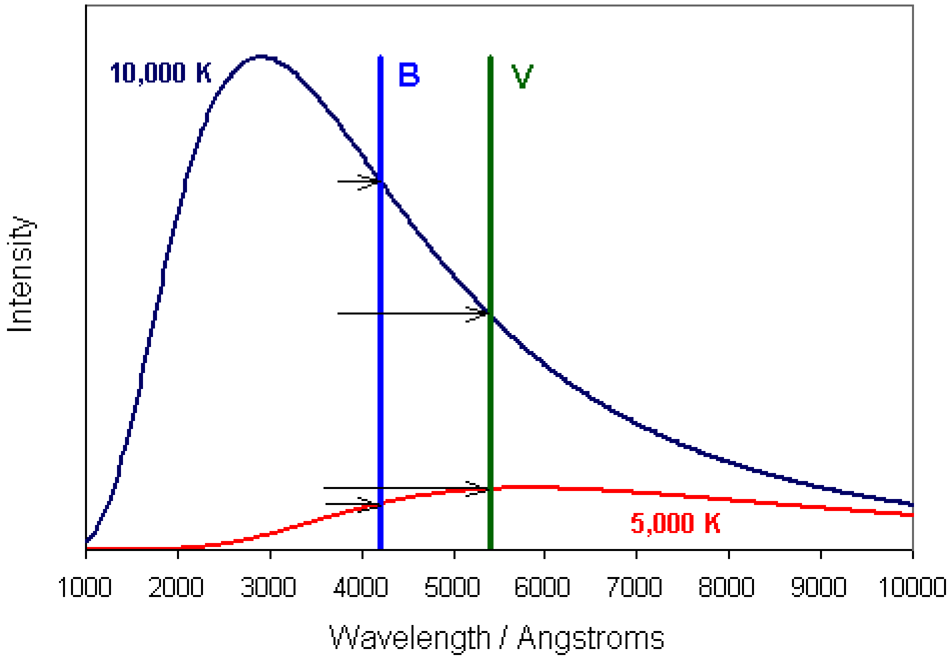


Figure 4. Comparison between the continuous spectrum of light emitted by two stars of different temperatures. The hotter star with temperature 10,000 K emits more light in the blue (short wavelength) part of the spectrum, the cooler star at 5,000 K is much redder. The diagram also depicts the central wavelengths of the B and V filters (see Figure 3). The arrows indicate the brightness of the stars in each of these filters, see text.

Interstellar Reddening

The space between the stars is not empty - it is filled with tenuous gas and clouds of dust that astronomers call the interstellar medium or ISM. The observed brightnesses and colours of stars are modified by the effect of the ISM through which the light passes on its way to the Earth. In general, the effects of absorption and scattering by dust in the ISM result in the light from the star appearing fainter and redder. For the typical sizes of dust grains in the ISM, this is referred to as Rayleigh scattering in which the angle through which the light is scattered is $\propto \lambda^{-4}$. Hence blue light (shorter wavelength) is scattered out of the beam and the object appears redder. This is the same effect that makes the sky appear blue.

The colour index $B-V$ that we measure after the light has been modified by the ISM is related to the intrinsic colour index $(B-V)_0$ that the star itself emitted via

$$B - V = (B-V)_0 + E(B-V) \qquad (10)$$

where $E(B-V)$ is the colour excess. Note that if the star is reddened, then this is a positive number.

For normal regions of the ISM the colour excess is directly related to the amount by which the star's brightness is reduced. In fact, the visual extinction A_V (the number of V magnitudes by which the star appears dimmer) is given by

$$A_V = (3.2 \pm 0.2) \times E(B-V)\ . \qquad (11)$$

Note that distance modulus is affected by interstellar extinction in the sense that because the star appears fainter we would think the star was farther away when in fact it is just being observed through a significant density of interstellar medium. Equation (6) for distance modulus is then modified to:

$$m_V - M_V = 5\log_{10} d - 5 + A_V \ . \qquad (12)$$

The amount of extinction varies as a function of wavelength such that it is highest at short wavelengths and lowest at longer wavelengths, see Figure 5. overall, extinction is roughly inversely proportional to wavelength. This is why extinction curves like that of Figure 5 are usually plotted with the inverse of wavelength (the so-called wavenumber) on the horizontal axis; in which case the curve is then flattened to a straight line at least through the visible part of the spectrum. Note the 2200 Angstrom bump in Figure 5 which is probably due to the presence of carbon in the form of spherical graphite grains in the ISM.

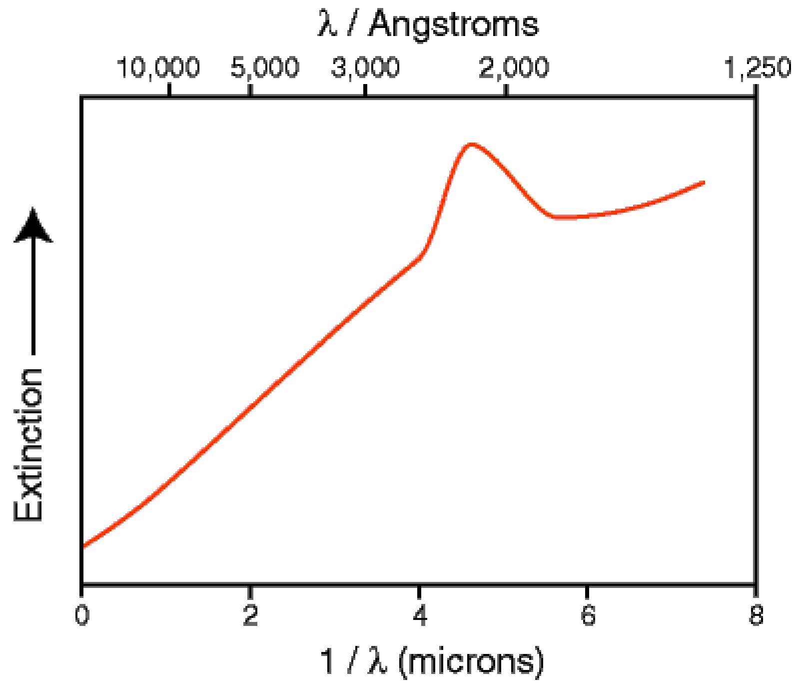


Figure 5. The average galactic extinction curve showing the increasing dimming effect of interstellar dust as wavelength decreases. The upper horizontal axis is just wavelength measured in angstroms whereas the lower horizontal axis is wavenumber with wavelength measured in microns (1 micron = 10,000 Angstroms).

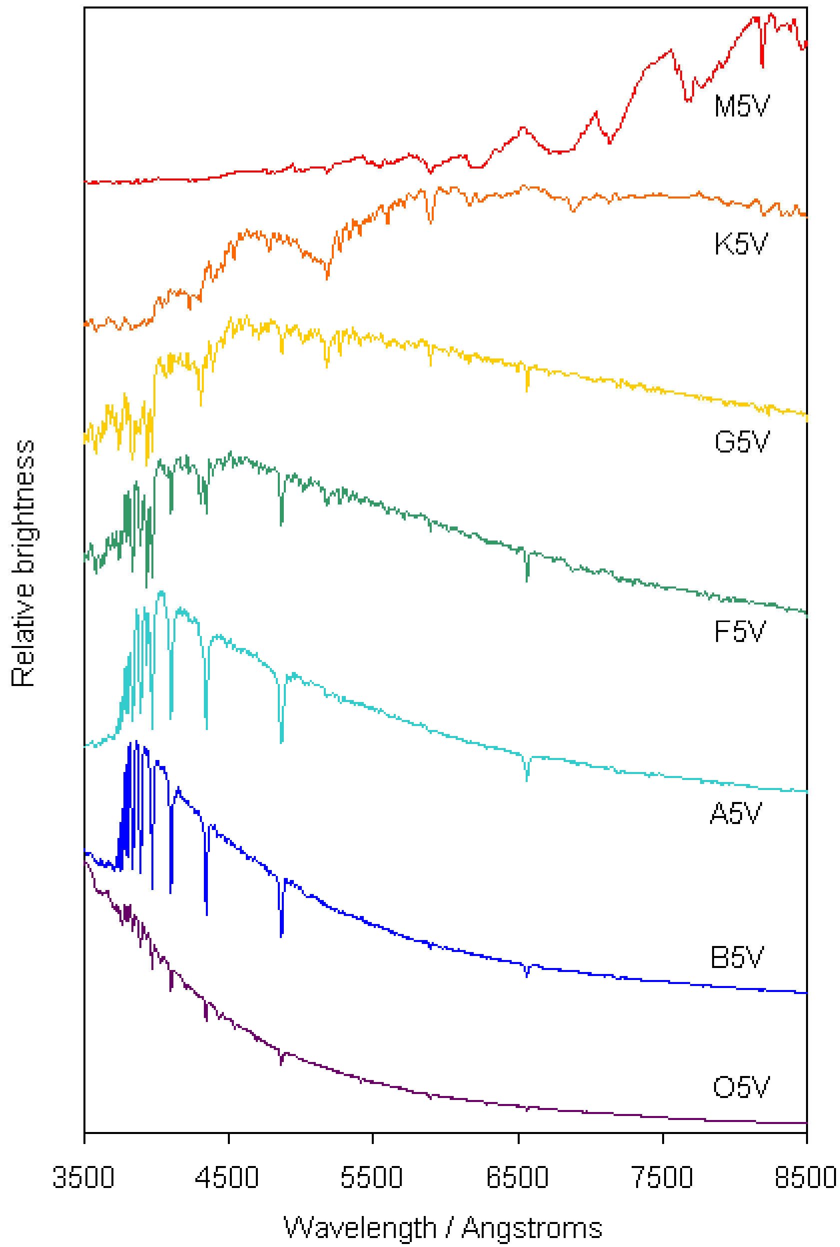
By what factor is the visual brightness of an object at the galactic centre reduced if it suffers from 30 magnitudes of visual extinction?

Do you think astronomers are more likely to investigate regions of high extinction, such as the galactic centre or star-forming regions, at infrared or optical wavelengths?

e spectral types of stars

Comparing the brightness of two stars using broadband filters like B and V gives us crude information on the colour of the star. More detailed investigations can be made by splitting the light from the star into a spectrum using a spectrometer (the light is dispersed using a prism or a diffraction grating - similar to the effect one sees when looking at light reflected from a compact disc). In fact, spectroscopy can provide astronomers with information about the chemical content, densities, temperatures and velocities of the material which is emitting the light.

When astronomers first collected spectra from a large number of stars it was realised that they could be classified according to their **spectral type**. In the standard MK system the spectral type is defined by a letter (O, B, A, F, G, K or M) and a number 0 to 9. The sequence then runs O0, O1, O2... O8, O9, B0, B1... and so on up to M9. Representative spectra for spectral types O5 through M5 are shown in Figure 6. You see that O stars are bluest and M stars are reddest. By analogy with the heating of solid objects which change colour from a dull red to blue-white as they increase in temperature we see that the spectral type is telling us about the temperature of the star (just as with colour index although with more detail). Note that the hotter stars are sometimes termed 'early-type' and the cooler ones 'late-type'.



M5V
K5V
G5V
F5V
A5V
B5V
O5V
Relative brightness
3500
4500
5500
6500
7500
8500
Wavelength / Angstroms

Figure 6. A compilation of actual spectra from stars of different spectral types. Each spectrum is shifted vertically by an arbitrary amount to separate it from its neighbours (data from Pickles 1998).

You might also notice in Figure 6 that the spectra are not smooth (in contrast to our simple examples of Figure 4). Rather they are punctuated by dips at particular wavelengths. For example, note the dip at about 4860 angstroms visible in several of the spectra and strongest in the A-type star. We have met this phenomenon several times before - when we discussed redshifts of galaxies and the abundances of chemical elements. You should remember that these dips are caused by absorption of the light from the hotter inner regions of the star's atmosphere by hydrogen gas in the cooler outer layers. These features are therefore referred to as absorption lines. Most stellar spectra show a large number of different absorption lines from a variety of chemical elements including some molecules such as water and titanium oxide. In fact, the classification into spectral types is made according to the relative strengths of various absorption lines and is effectively a sequence of temperatures, see Table 3 (note 'metals' is the term used by astronomers to refer to all elements other than hydrogen and helium!).

Table 3. The MK spectral classification scheme.

Spectral type	Temperature / K	Colour	Dominant absorption lines
O	>20,000	Hottest blue stars	Ionized helium, strong ultraviolet continuum
B	20-10,000	Hot blue stars	Neutral helium dominates (He I)
A	10-7,000	Blue/blue-white stars	Neutral hydrogen (H) dominates
F	7-6,000	White stars	Singly ionized calcium (CaII), neutral H weaker, other metals
G	6-5,000	Yellow stars	Ca II dominates, neutral metals (e.g. iron - FeI)
K	5,000-3,500	Orange-red stars	Neutral metals (Ca, Fe) dominate, molecular bands appear
M	3,500-2,000	Coolest red stars	Molecular bands dominate (e.g. titanium oxide - TiO), neutral metals

You may have noticed that the spectral types in Table 1 and labelled in Figure 6 have a letter V attached to the end. In fact, this is not formally part of the spectral type but is a second parameter called the **luminosity class** defined by the ratio of strengths of various pairs of absorption lines. It is written as a roman numeral in the range I to VII. Stars like the Sun are luminosity class V. In effect this is a sequence of density in the stellar atmosphere - stars which are brighter for the same effective temperature must be larger and therefore have more tenuous, less dense, outer layers. They follow the scheme given in Table 4.

Table 4. Luminosity classes of stars.

Ia	Ib	II	III	IV	V	VI	VII
Brightest	...	...	...	...	...	...	Faintest
Lowest density	...	...	...	...	...	...	Highest density
Bright supergiants	Supergiants	Bright giants	Giants	Subgiants	Main sequence	Sub-dwarfs	White dwarfs

The sun is of type G2V with effective temperature 5800K.

The Hertzsprung-Russell diagram

In 1911, Ejnar Hertzprung plotted the first diagram of the relative magnitudes of stars in a cluster versus their spectral types. Two years later Henry Russell, working independently, produced a plot of the absolute magnitude of nearby stars with well-determined distances against their spectral types. The resulting diagram has come to be known as the Hertzsprung-Russell or HR diagram. It is one of the most useful diagrams in astronomy and is fundamental to our understanding of stars.

The HR diagram is plotted with intrinsic brightness (or any other equivalent quantity e.g. luminosity or absolute magnitude) on the vertical axis, increasing upwards, and effective temperature (or colour index or spectral type) on the horizontal axis, increasing to the left. The most common HR diagrams are called colour-magnitude diagrams after the quantities which are plotted. Figure 7 is an HR diagram constructed from data extracted from the catalogue made using observations with the Hipparcos satellite.

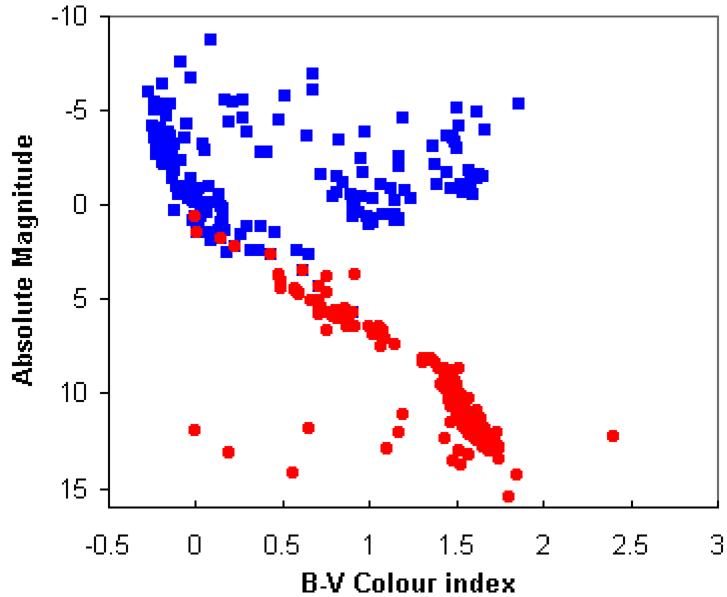


Figure 7(a). A composite HR diagram containing two datasets - all the stars closer than 10 parsecs (red circles) and all the stars with apparent magnitudes brighter than 3 (blue squares) - data from the Hipparcos catalogue. Here account has been taken of each star's distance so that brightness can be measured in *absolute* visual magnitudes. [Note this diagram gives a false representation of a typical HR diagram as the top half is artificially over-populated by forcibly including the bright stars - as we discussed earlier most stars are actually fainter than the Sun.]

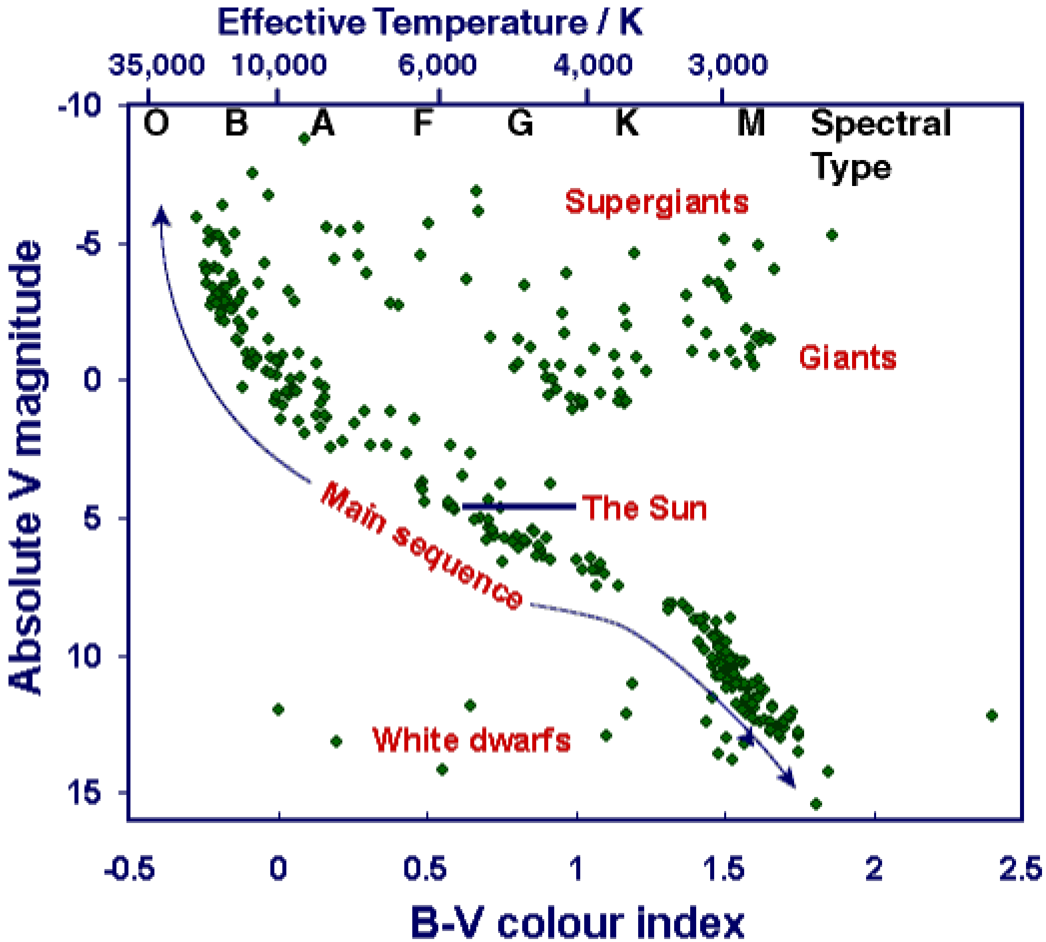


Figure 7. Same as (a) except here the lower horizontal axis is labelled in B-V colour index and the upper axis with spectral type and effective stellar temperature in degrees Kelvin. A number of luminosity classes are also marked: VII, the faintest (white dwarfs); V (main sequence); III (giants); and I, the brightest (supergiants).

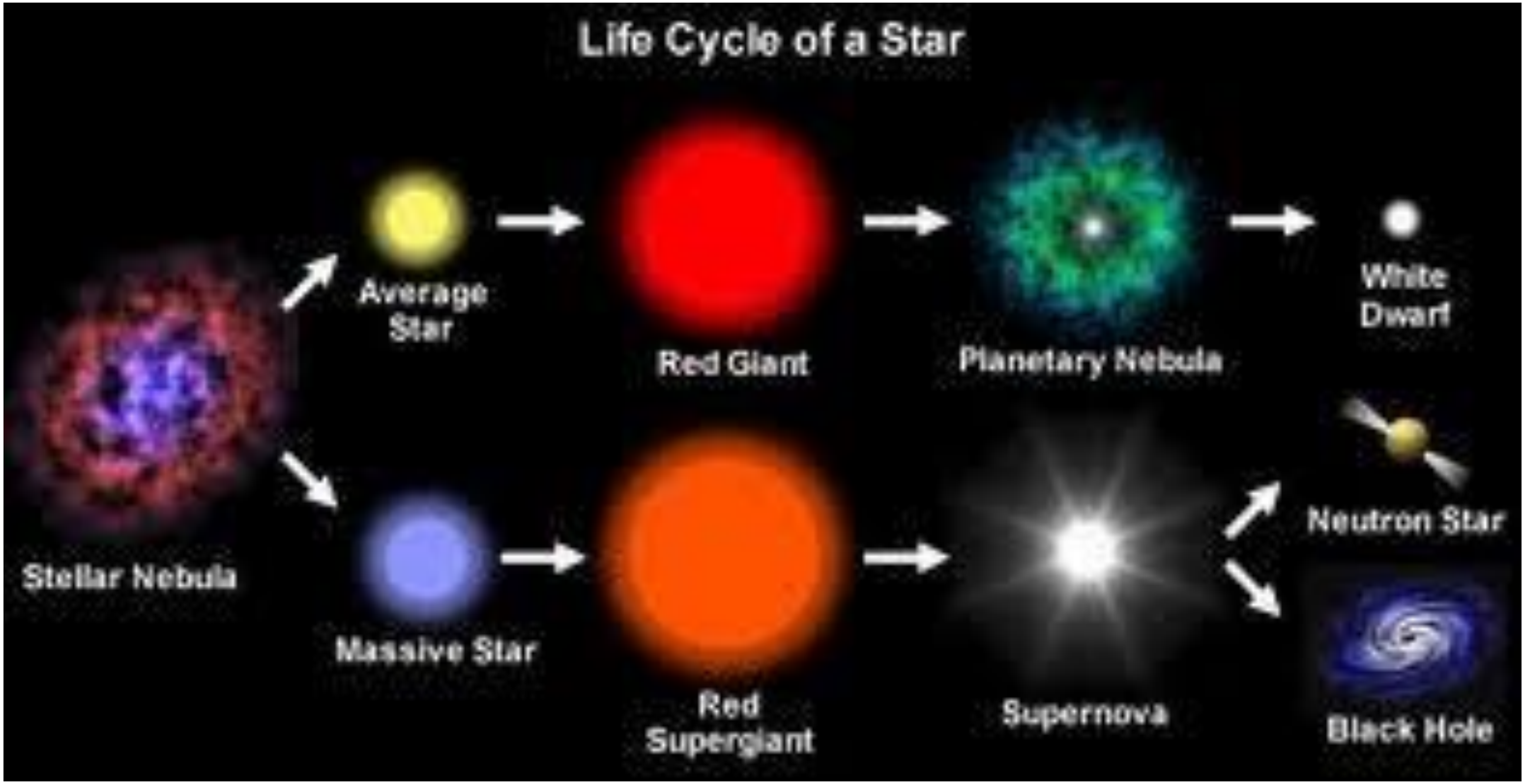
The main points to note from Figure 7 are that 90% of stars lie on a narrow diagonal band running from top left (bright and hot) to bottom right (faint and cool). This is called the main sequence. The sun lies approximately in the middle of the main sequence. Giants and supergiants are much more luminous for a given temperature, lying above the main sequence and generally at the cooler end (the right) of the HR diagram. Far below the main sequence are the faint white dwarf stars.

As we shall see in the next few modules when we consider the sun and the evolution of stars the HR diagram is an invaluable tool in modern astronomy.

nclusions

You should now understand the astronomical magnitude system for measuring brightnesses, the characterisation of colour by the colour index and the spectral classification of stars via spectral types and luminosity classes. We will go on in the next module to look at a typical main sequence star, our own Sun, and investigate in particular its source of energy.

LifeCyclesofStars


Life Cycle of a Star
Average Star
Red Giant
Planetary Nebula
White Dwarf
Stellar Nebula
Massive Star
Red Supergiant
Supernova
Neutron Star
Black Hole

Theuniversestartedwiththe…

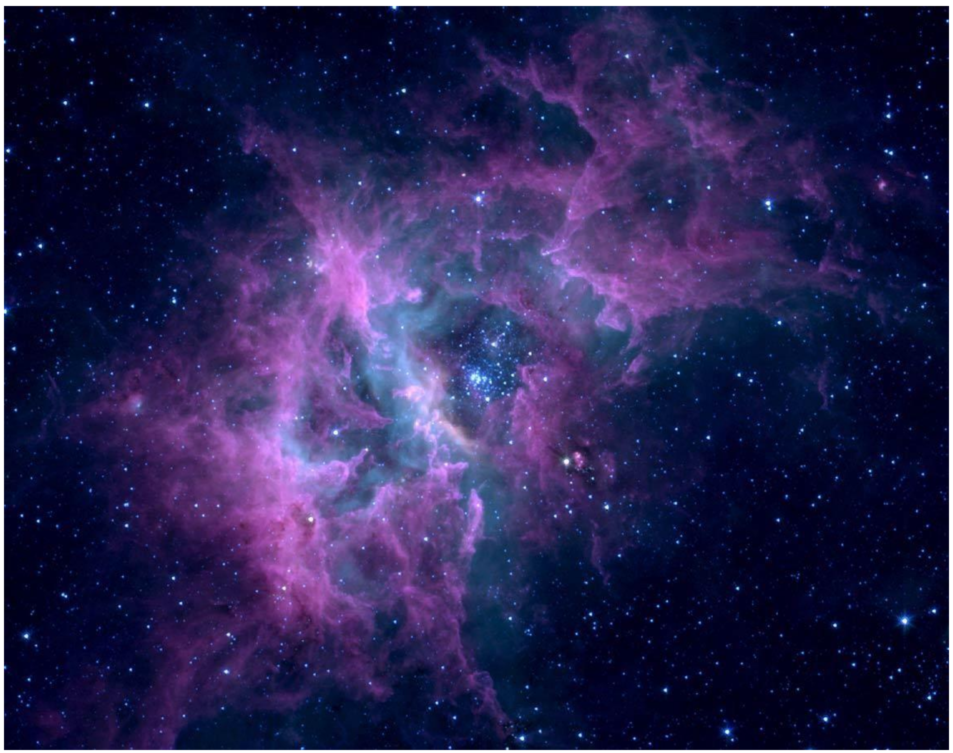
Big Bang!

- Everything continued to expand, clouds of dust started to gravitate towards each other forming stars.

Astar'slifebeginsina…

Nebula!

- A cloud of gas and dust, consisting mostly of Hydrogen



Astar'slifebegins…

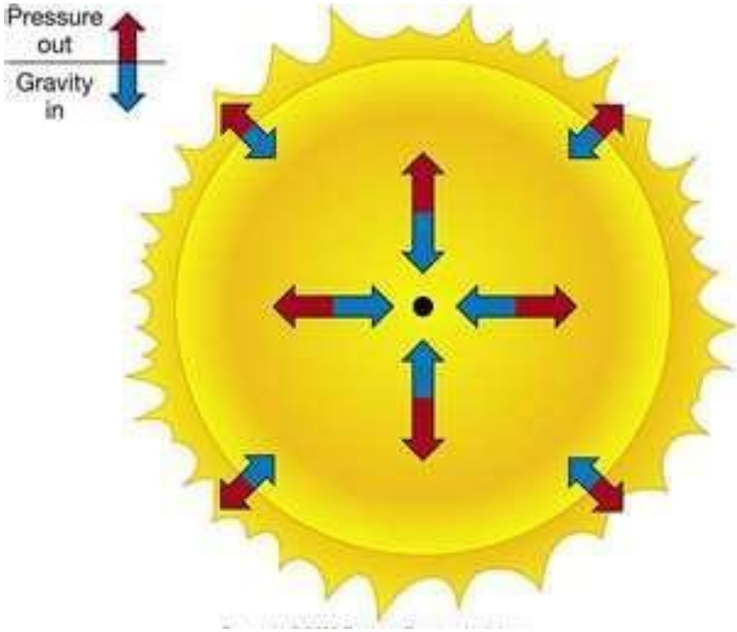
- Gas and dust begin to clump together to form a **Protostar** (a babystar).

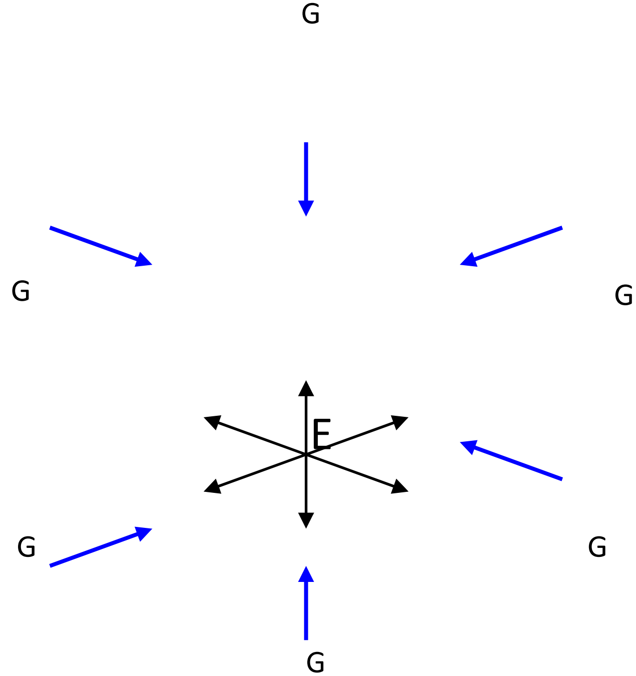
Astar’slifebegins…

- The smaller a star is the longer it will live.
  - Larger stars have more fuel, but they have to burn (fuse) it faster in order to maintainequilibrium.
  - Because fusion occurs at a faster rate in massive stars, large stars use all their fuel in a shorter length of time.
  - So…Asmallerstarhaslessfuel,butitsrateof fusion is not as fast.Therefore, smaller stars live longer than larger stars because their rate of fuel consumption is not as rapid.

Astar'slifebegins…

- The star's main goal in life is to achieve stability, or <u>equilibrium,</u> where pressure from fusion within the core is equal to the force of gravity pushing down on it (this keeps the star "alive").





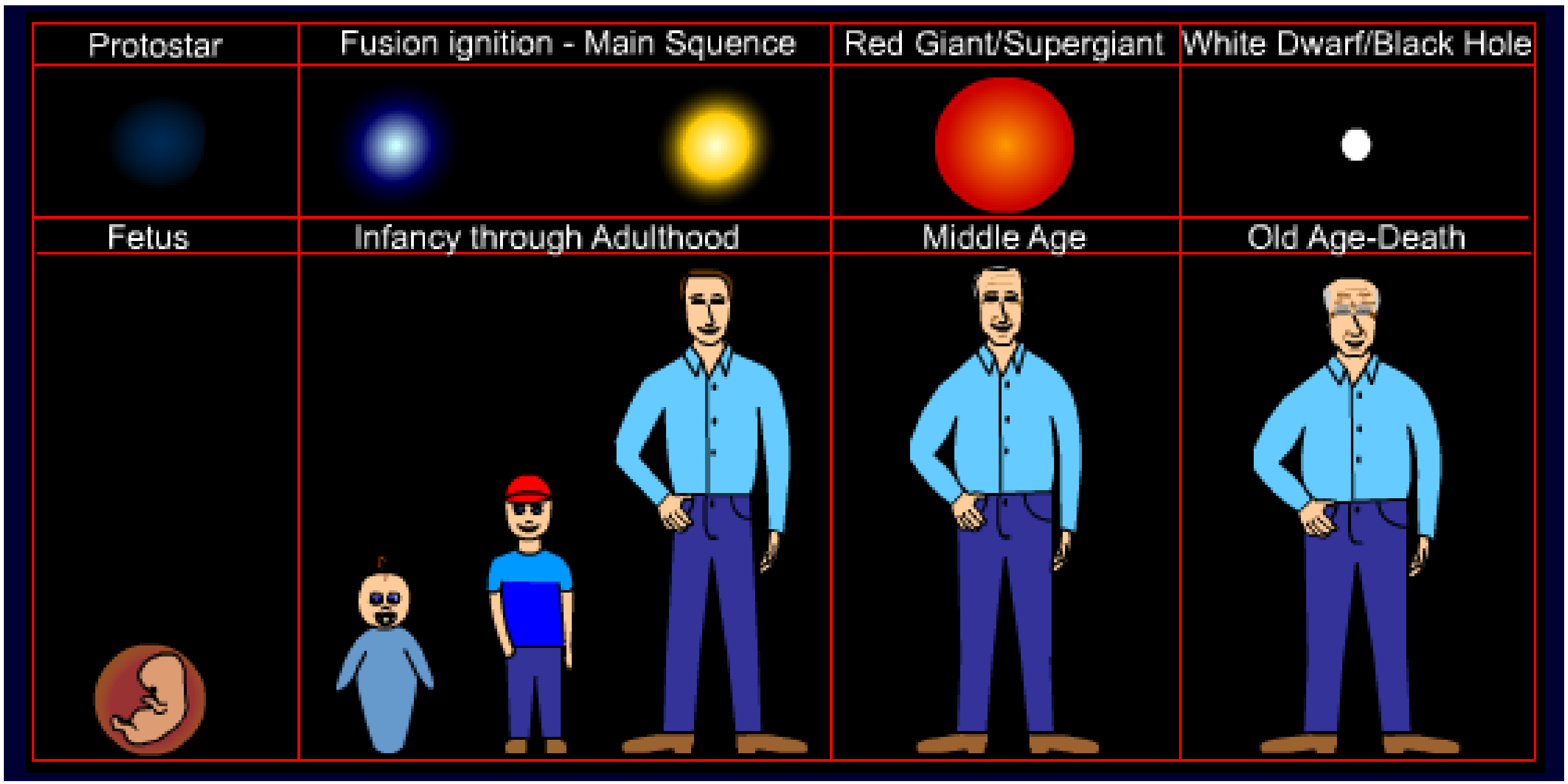
Astar'slifebegins…

Continuous steps occur inside the core of a main sequence star, until there is no more Hydrogen.

- Step 1 - Nuclear fusion (hydrogen turning to helium). Gravity = gas pressure (equilibrium)
- Step 2 - Out of fuel
- Step 3 - Fusion stops, temperature drops
- Step 4 - Core contracts (gravity pulling atoms in)
- Step 5 - Increased temperature (more atoms, more collisions) and density in the core reinitiates nuclear fusion, equilibrium is achieved, and the cycle beginsagain at Step 1.

LifeCycleofaStarlikeourSun…

Nebula Protostar Main Sequence Star Red Giant Planetary
Nebula White Dwarf



LifeCycleofaStarlikeourSun…

- Our sun is at the Main Sequence stage in its life.
  - When the hydrogen in the core has been used up, the core shrinks and hydrogen fusion begins in the outer layers,
  - which then expands the entire star, turning it into a Red Giant.
  - The sun begins to die when helium is fusing intoother elements,thenthegasesatthesun'ssurfacestartto blow away in bursts, called a Planetary Nebula (or halo of gases,
  - Resulting in a hot carbon-oxygen core called a White Dwarf.

LifeCycleofaStarWithGreaterMass

Than OurSun…

Nebula

Protostar

Main Sequence Star Red Supergiant

Supernova

Black Hole

or

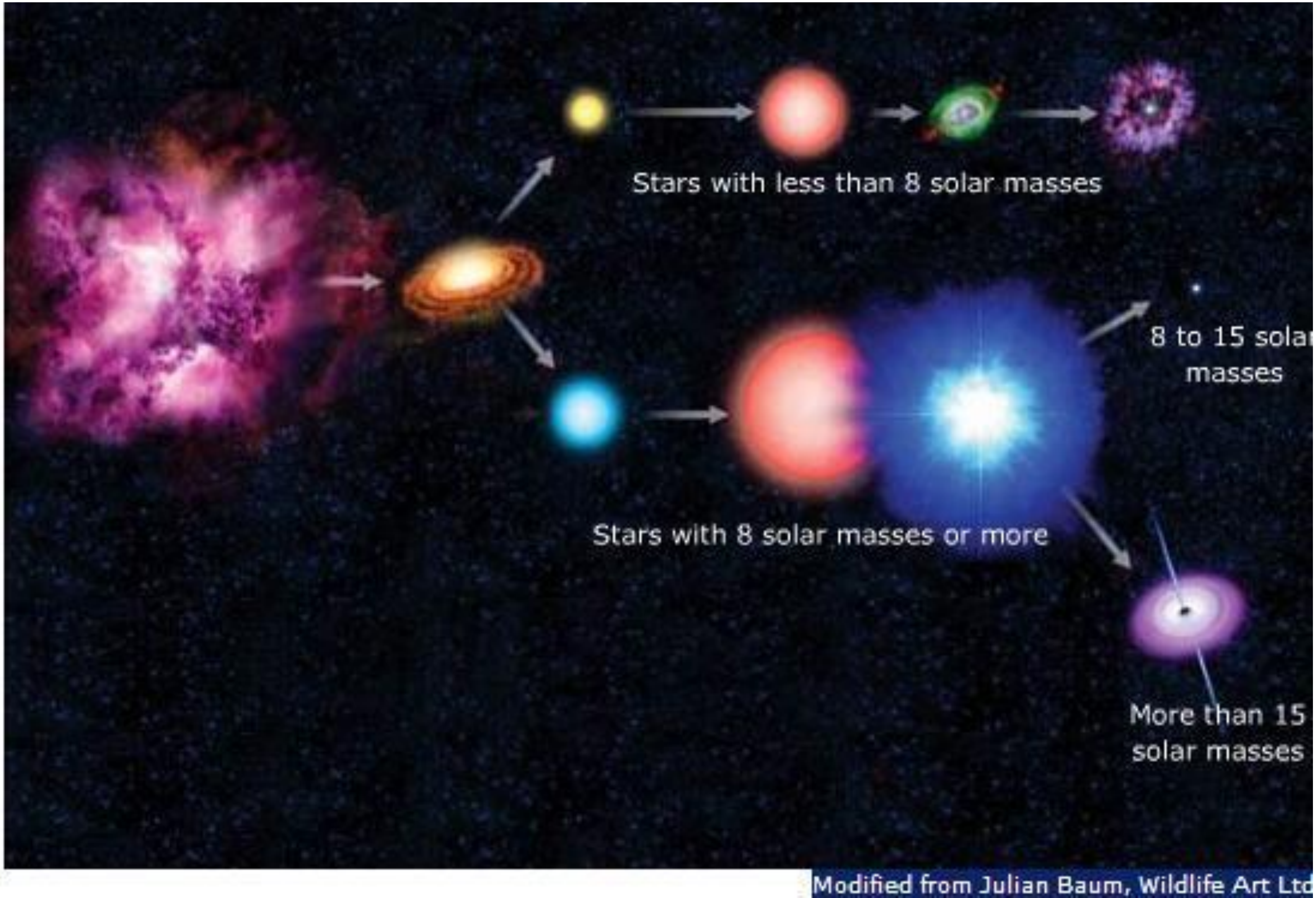
Neutron Star

LifeCycleofaStarWithGreaterMass

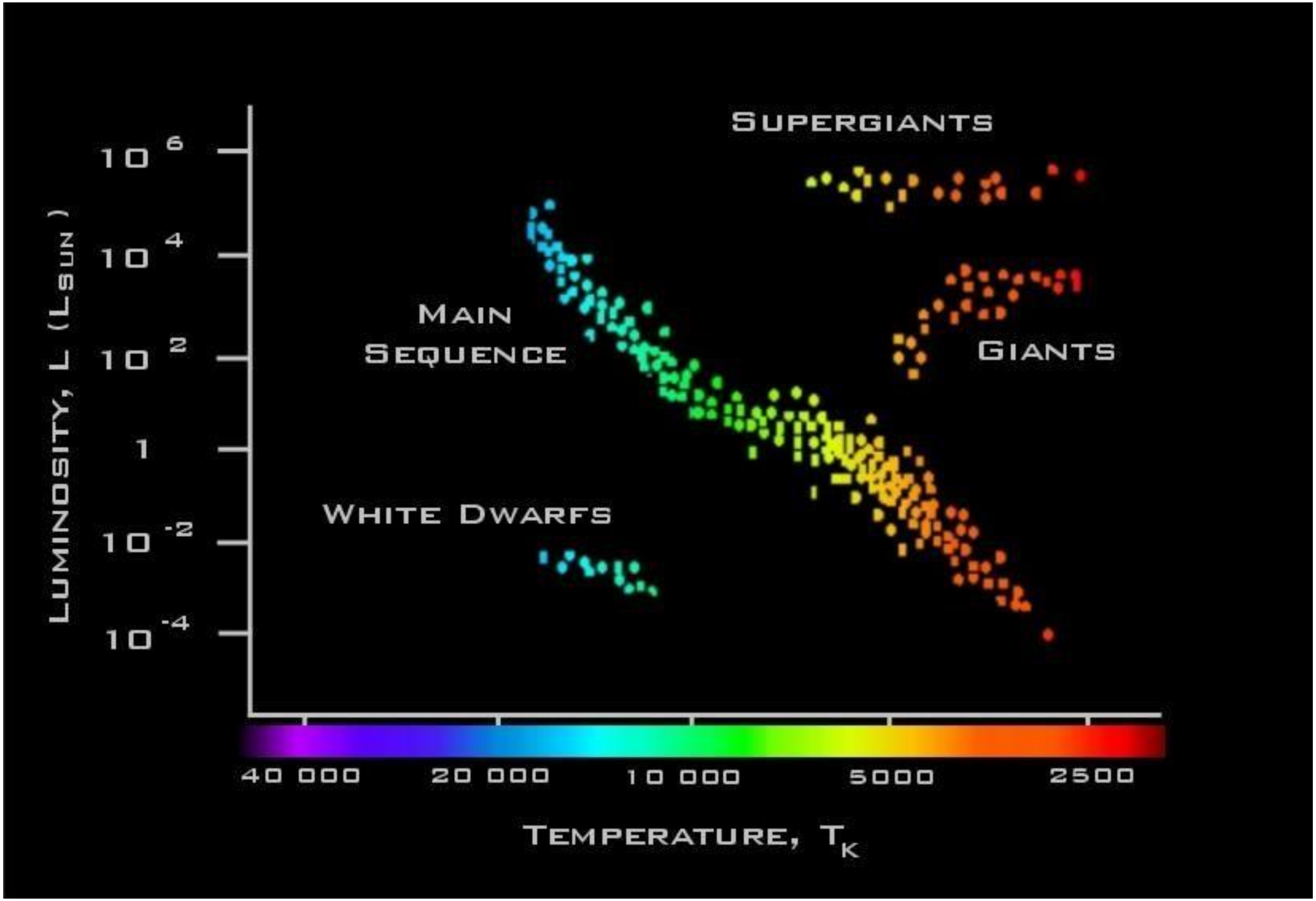
Than OurSun…

- Massive stars go through the same life stages as our sun (just on a larger scale) upto the Main Sequence stage,
- Then the massive stars expand into a Red Supergiant,
- Explode into a Supernova,
- Then turn into a Black Hole or a Neutron Star.

Life Cycle of Stars



Hertzsprung-Russell (HR) Diagram

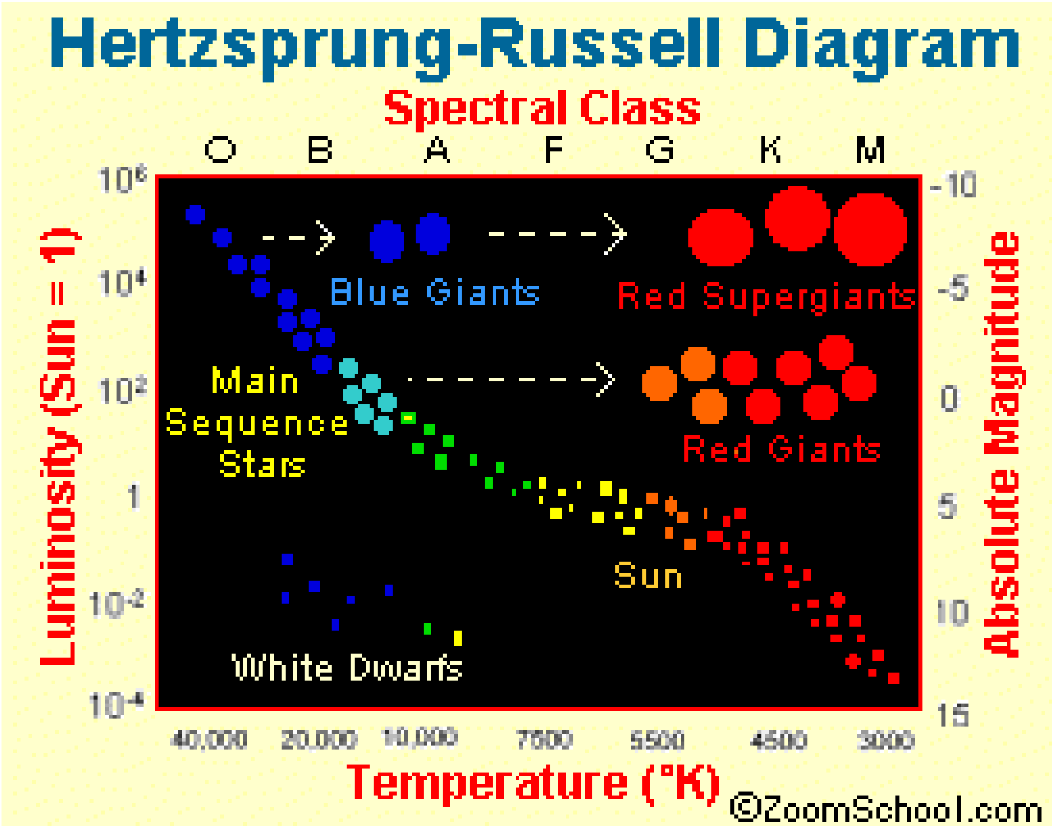


Supergiants
Main Sequence
Giants
White Dwarfs
Luminosity, L (Lsun)
10^6
10^4
10^2
1
10^{-2}
10^{-4}
40 000
20 000
10 000
5000
2500
Temperature, T_K

HRDiagram

- The Hertzsprung-Russell (HR) Diagram is a tool that shows relationships and differences between stars (temperatures, brightness, colors, etc.)
- It is something of a "family portrait." It shows stars of different ages and in different stages, all at the same time.
  - A star in the upper left corner of the diagram would be hotand bright.
  - A star in the upper right corner of the diagram would be cool and

    bright.
  - The Sun rests approximately in the middle of the diagram, and it is the star which we use for comparison.
  - A star in the lower left corner of the diagram would be hotand dim.
  - A star in the lower right corner of the diagram would be coldand dim.



UNIT 2

INDIA’S SPACE PROGRAMME

7. Economic Aspects of India’s Space Program

By the early nineties, all the four major components of the space programe, namely, Satellite Communications, Meteorology, Earth Observations and Launch Vehicles had entered the operational stage.

The Satellite-based Communication Services (SATCOM) Policy of 1997 and the Remote Sensing based value added services envisaged opening of the space industry to the private sector. Therefore it was considered timely and appropriate to commission a study on the economic aspects of the Indian space program through the Madras School of Economics (Sankar 2006a; 2006b; Sankar et al. 2003).

7.1. Space Expenditures

Accumulated space expenditures since inception to the last fiscal year ending on March 31, 2006 amounted to US$ 7 billion.

As is obvious from the figure, 39 per cent investment is on launch vehicles, 36 per cent on Satellite communications and meteorology, 14 per cent on earth observations, 6 per cent on space sciences and the balance on other items. About three-fourth of the total expenditure was incurred towards development of technology in the case of launch vehicles, whereas in the case of satellite communications, meteorology and earth observations, three-fourth of the investment is for building operational systems based on service needs of the country.

The space expenditure of India as a percentage of gross domestic product (GDP) today stands at 0.09 per cent. Compared with the current annual government space budgets of US $2.5 billion for Japan, and US$1.5 billion for France, India's space budget is US$0.60 billion.

7.2. Methodology

For the purposes of economic analysis, it is useful to classify space activities into two stages, namely, (i) design, development, testing, manufacturing and launch of spacecrafts into desired orbital slots (construction stage), and (ii) applications of satellite services to different uses (exploitation stage). The output basket of the space program contains a mix of private goods, public goods, social goods and strategic/incommensurable goods. Research in space sciences, most meteorological services and information are public goods. Equity considerations are important in provision of certain goods e.g. access to public telephone, access to radio and TV. The social goals dominate in public sector radio and TV programs. Use of the space program as an instrument for guaranteeing strategic, political, scientific and economic leadership yields strategic and incommensurable benefits.

7.3. Construction Stage

Regardless of the nature of goods / services provided and whether it is produced by a public firm or private firm, cost minimization is a valid criterion. The economic costing methodology requires (i) a rational basis for allocation of costs among the payloads of a multi-purpose satellite (ii) apportionment of common and joint costs amongst various ongoing programs of the organization / institution (iii) investment expenditure, their time pattern and cost of capital and (iv) output streams, their time pattern and discount rates for present value.

The global market for communication transponders is generally competitive with many private and public suppliers and many customers buying the transponders. Government induced market distortions are relatively less in this market. Hence, the international market prices can serve as a benchmark for assessing the cost effectiveness of INSAT transponders. A detailed study on economic costing of INSAT transponders with 10 per cent cost of capital on investments and 5.5 per cent discount factor on future returns has brought out the cost advantage of INSAT transponders by at least 25 per cent of the prevailing international prices. The cost performance of INSAT system has been considered to be commendable keeping in view the relatively high capital cost in India and the dependence on some foreign components in the production of the satellites.

A comparative analysis of remote sensing satellites and launch vehicles is rather difficult due to non-availability of reliable estimates of the costs of foreign systems and also due to differences in capabilities. However, preliminary estimates show that the costs of Indian Remote Sensing Satellite (IRS-1D) is very much lower than the reported costs of similar LANDSAT and SPOT satellites. Similarly, the development cost of India’s PSLV and GSLV is US $1.3 billion as compared to about US $ 4 billion for the European Ariane 1 to 4, though there are some capacity variations in these systems.

7.4. Exploitation Stage

For measurement of the benefits, the role of satellite technology is considered under three different categories: (a) where the technology is unique, (b) where the technology is a substitute to existing technologies, and (c) where the technology is complementary to existing technologies. In the second case, one can measure cost savings due to satellite technology compared with the existing technology. If the technology is superior to the existing one, one has to estimate the incremental value of the improvement. Where the space technology is used in conjunction with many other technologies, one has to rely on a cost allocation procedure or a benefit sharing method or on expert opinion to estimate the benefit attributable to the space technology.

The INSAT system has played a key role in augmenting Broadcasting, Telecommunications and Meteorological services in the country and has contributed immensely to economic and social development. Satellite communication technologies are terrain and distance independent and they enable governments to achieve goals such as the development of backward and remote areas at low costs and in a short time and thereby achieve technological leapfrogging.

7.4.1. Television

The Major benefits of the INSAT system to Doordarshan (public TV) are expansion in area coverage from 14 per cent in 1983 to 78 per cent in 2005, population coverage from 26 per cent in 1983 to 90 per cent in 2005, increase in the number of channels from 2 to 32, remote area coverage, satellite news gathering, dissemination of weather and cyclone warning and use of TV as a media for training and education.

A detailed analysis show that for enhancing the population coverage further from 90 to 100 per cent with the distribution of a bouquet of 20 DD channels by the public broadcaster Doordarshan, the capital cost and annual operating cost through terrestrial technology is Rs.34560 million and Rs.5184 million respectively while a satellite based solution with direct reception at homes, would involve a capital cost of Rs.6380 million and annual operating cost of Rs.357 million. Thus, given the unique physiographical feature of India, the satellite communications is the least-cost option for achieving 100 per cent population coverage.

The growth of satellite TV has also aided in the emergence of new economic activities. The advent of satellite TV contributed to the growth of several industries like the manufacturing of TV sets, cables, receiving antenna and other equipment and program production. There are about 100,000 cable TV operators and about 35 million cable TV households in the country. The gross earnings of cable TV operators is nearing Rs. 10 billion.

7.4.2. Telecommunications

Remote area communication is an important objective of public policy. There is considerable cost savings due to use of satellite technology compared with the alternative of optical fiber cable network in remote area communication. The cost of connecting 393 remote areas, currently served by INSAT, by optical fibre cable would be Rs.23580 million while the comparable cost for satellite technology would be Rs.10460 million. It may be noted that there are 30,000 remote villages of similar nature needing connectivity. The other uses of satellite technology are: alternative media back up for terrestrial services, business communications, portable terminals for disaster management, Tele-medicine and Satellite Aided Search and Rescue.

Apart from the cost saving, there are many external benefits which are diffused economy-wide. In case of Andaman and Nicobar (AN), rapid expansion of telecom since the mid-nineties facilitated the integration of AN with the mainland thereby boosting the growth of industry, trade and tourism and raising the growth rate of gross state domestic product to more than 8 per cent.

7.4.3. Meterology

Satellites have made significant contributions to the generation of meteorological information by extending observation to oceans and remote areas on land, enabling generation of new types of observations, facilitating new concepts of data assimilation into models, reducing costs of a few types of observations and enhancing the reliability of certain types of data.

Meteorological services are recognized as public goods. The major contributions of satellite technology are in the areas of weather technology (cloud motion vector, wind-sea surface temperature and outgoing long wave radiation) and tropical cyclone (identification of genesis and current position, intensity of change and transmission of cyclone warnings). A comparative study of 1977 (before INSAT) and 1990 (after INSAT) cyclones which hit Andhra Pradesh, shows that even though the two cyclones are similar, due to the successful tracking of the cyclone in 1990 with the INSAT imaging instrument (VHRR) and the success of preparatory steps taken by the government, the loss of lives in 1990 was only 817 compared with 10,000 in 1977. This is an important incommensurable benefit of satellite technology.

7.4.4. Remote Sensing

The advantages of remote sensing are synoptic coverage, multi-spectral capability, multi-temporal capability and digital capture of data. Remote sensing technology is being used in three different situations. It is an exclusive tool for estimation of snow melt run-off, rapid assessment of areas affected by natural disasters, identification of potential fishing zones in offshore areas and mapping of inaccessible areas. It is a substitute tool to conventional methods in mapping of land use, waste lands, and urban land use; preparing ground water prospect maps, watershed development plan, coastal zone management plan etc; and in monitoring forest cover, urban sprawl, status of environment etc. It is a complement in cases like area and crop forecasting and urban development plans. Its advantage is that it yields unbiased, timely and enhanced information. Based on case studies of applications of remote sensing in India’s development programs, Table 3 provides estimates of investments, direct returns, and economic benefits.

Apart from the major benefits enumerated above, the policy of self-reliance has also enabled internal competence building and technology development and spin-offs to non-space sectors. For example, the spin-off outputs till 2005 include 224 Technology Transfers, 165 patents, 10 trademarks and 17 copyrights. ISRO has nurtured a symbiotic partnership with more than 500 Indian firms. The flow of funds to industry currently is about 40 per cent of the space budget. This partnership has generated significant spin-off effects to the industries in terms of improved manufacturing processes, quality control and management practices.

Table 3: Investments and Benefits in Remote Sensing

A	*Investments*		Rs. Millions	
	Operational Missions		10,080	
	Data Reception, Processing and Applications		5,540	
B	*Direct Returns*			
1.	Returns from sale of Satellite Data and Value Added Products by NDC		1,600	
2.	Returns from ANTRIX through access fees and royalty		600	
3.	Opportunity cost (cost of foreign satellite data equivalent to IRS data used).		~ 5,000	
4.	Cost saving due to value addition		~ 12,000	
5.	Cost saving due to mapping using RS data		~ 11,000	
C	*Economic Benefits*		*Rs. Millions*	
	Program	Nature of Benefit	Estimate from Case Studies	Potential Benefit to the country in the Long-run
1.	National Drinking Water Technology Mission	Cost saving due to increase in success rate	2,560 (5 States)	5,000 – 8,000
2.	Urban Area Perspective / Development / Zonal / Amenities Plan for Cities / Towns	Cost saving in mapping	50.4 (6 Cities)	16,000 – 20,000

3.	Forest Working Plan	Cost saving in mapping	2,000 (200 Divisions)	11,860
4.	Potential Fishing Zone Advisories	Cost saving due to avoidance of trips in non-PFZ advisories	5,450	16,350
5.	Wasteland Mapping: Solid Land Reclamation	Productivity gain	990 (UP)	24,690
6.	Integrated Mission for Sustainable Development: Horticultural Development in Land With and Without Shrub	Gross income	Rs.0.20 to 0.40 (per hectare)	13,000 – 26,000
7.	Bio-prospecting for Medicinal Herbs	Value of Indian life saving drugs		800

Unit-I : The Cellular Concept-System Design Fundamentals

Introduction

We have seen that the technique of substituting a single high power transmitter by several low power transmitters to support many users is the backbone of the cellular concept. In practice, the following four parameters are most important while considering the cellular issues: system capacity, quality of service, spectrum efficiency and power management. Starting from the basic notion of a cell, we would deal with these parameters in the context of cellular engineering in this chapter.

What is a Cell?

The power of the radio signals transmitted by the BS decay as the signals travel away from it. A minimum amount of signal strength (let us say, x dB) is needed in order to be detected by the MS or mobile sets which may the hand-held personal units or those installed in the vehicles. The region over which the signal strength lies above this threshold value x dB is known as the coverage area of a BS and it must be a circular region, considering the BS to be isotropic radiator. Such a circle, which gives this actual radio coverage, is called the foot print of a cell (in reality, it is amorphous). It might so happen that either there may be an overlap between any two such side by side circles or there might be a gap between the

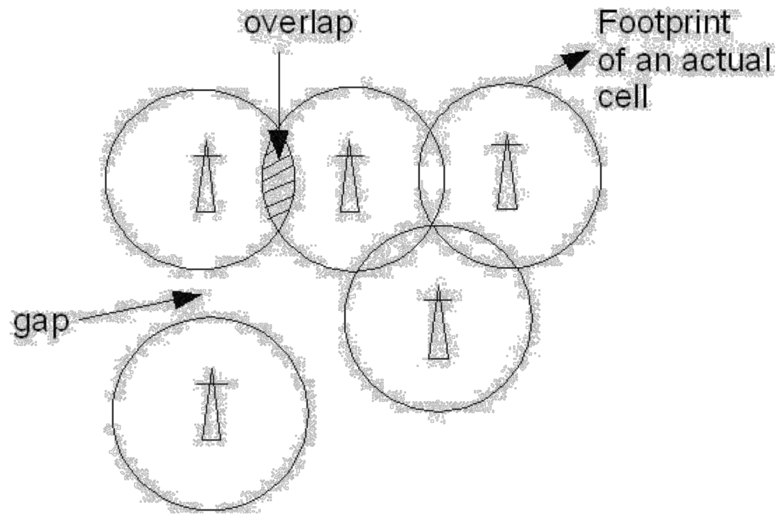


Figure 3.1: Footprint of cells showing the overlaps and gaps.

coverage areas of two adjacent circles. This is shown in Figure 3.1. Such a circular geometry, therefore, cannot serve as a regular shape to describe cells. We need a regular shape for cellular design over a territory which can be served by 3 regular polygons, namely, equilateral triangle, square and regular hexagon, which can cover the entire area without any overlap and gaps. Along with its regularity, a cell must be designed such that it is most reliable too, i.e., it supports even the weakest mobile with occurs at the edges of the cell. For any distance between the center and the farthest point in the cell from it, a regular hexagon covers the maximum area. Hence regular hexagonal geometry is used as the cells in mobile communication.

Frequency Reuse

Frequency reuse, or, frequency planning, is a technique of reusing frequencies and channels within a communication system to improve capacity and spectral efficiency. Frequency reuse is one of the fundamental concepts on which commercial wireless systems are based that involve the partitioning of an RF radiating area into cells. The increased capacity in a commercial wireless network, compared with a network with a single transmitter, comes from the fact that the same radio frequency can be reused in a different area for a completely different transmission.

Frequency reuse in mobile cellular systems means that frequencies allocated to

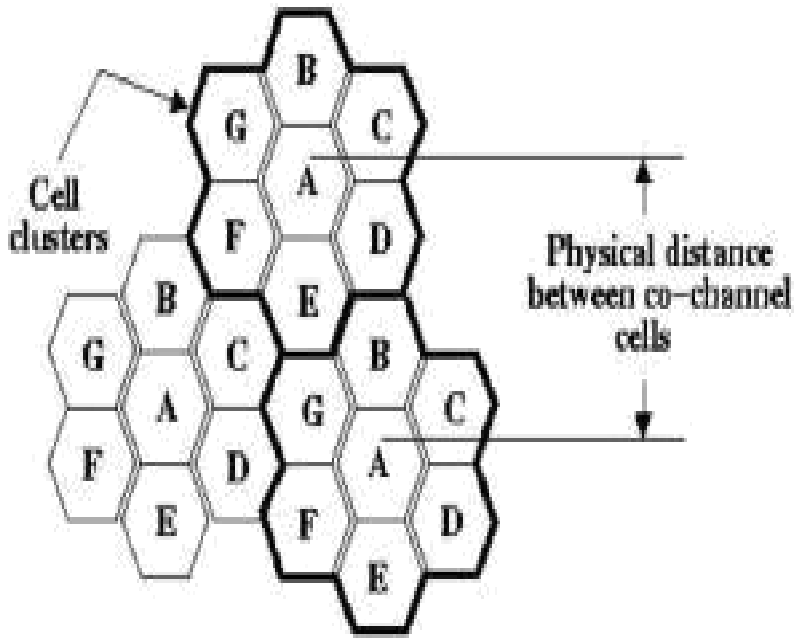


Figure 3.2: Frequency reuse technique of a cellular system

the service reused in a regular pattern of cells, each covered by one base station. The repeating regular pattern of cells is called cluster. Since each cell is designed to use radio frequencies only within its boundaries, the same frequencies can be reused in other cells not far away without interference, in another cluster. Such cells are called ‘co-channel’ cells. The reuse of frequencies enables a cellular system to handle a huge number of calls with a limited number of channels. Figure 3.2 shows a frequency planning with cluster size of 7, showing the co-channels cells in different clusters by the same letter. The closest distance between the co-channel cells (in different clusters) is determined by the choice of the cluster size and the layout of the cell cluster. Consider a cellular system with S duplex channels available for use and let N be the number of cells in a cluster. If each cell is allotted K duplex channels with all being allotted unique and disjoint channel groups we have $S = KN$ under normal circumstances. Now, if the cluster are repeated M times within the total area, the total number of duplex channels, or, the total number of users in the

system would be $T = MS = KMN$. Clearly, if K and N remain constant, then

$$T \propto M \tag{3.1}$$

and, if T and K remain constant, then

$$N \propto \frac{1}{M} \tag{3.2}$$

Hence the capacity gain achieved is directly proportional to the number of times a cluster is repeated, as shown in (3.1), as well as, for a fixed cell size, small N decreases the size of the cluster with in turn results in the increase of the number of clusters (3.2) and hence the capacity. However for small N, co-channel cells are located much closer and hence more interference. The value of N is determined by calculating the amount of interference that can be tolerated for a sufficient quality communication. Hence the smallest N having interference below the tolerated limit is used. However, the cluster size N cannot take on any value and is given only by the following equation

$$N = i^2 + ij + j^2, \qquad i \geq 0, j \geq 0, \tag{3.3}$$

where i and j are integer numbers.

Channel Assignment Strategies

With the rapid increase in number of mobile users, the mobile service providers had to follow strategies which ensure the effective utilization of the limited radio spectrum. With increased capacity and low interference being the prime objectives, a frequency reuse scheme was helpful in achieving this objective . A variety of channel assignment strategies have been followed to aid these objectives. Channel assignment strategies are classified into two types: fixed and dynamic, as discussed below.

Fixed Channel Assignment (FCA)

In fixed channel assignment strategy each cell is allocated a fixed number of voice channels. Any communication within the cell can only be made with the designated unused channels of that particular cell. Suppose if all the channels are occupied, then the call is blocked and subscriber has to wait. This is simplest of the channel assignment strategies as it requires very simple circuitry but provides worst channel utilization. Later there was another approach in which the channels were borrowed from adjacent cell if all of its own designated channels were occupied. This was named as *borrowing strategy*. In such cases the MSC supervises the borrowing pro-cess and ensures that none of the calls in progress are interrupted.

Dynamic Channel Assignment (DCA)

In dynamic channel assignment strategy channels are temporarily assigned for use in cells for the duration of the call. Each time a call attempt is made from a cell the corresponding BS requests a channel from MSC. The MSC then allocates a channel to the requesting the BS. After the call is over the channel is returned and kept in a central pool. To avoid co-channel interference any channel that in use in one cell can only be reassigned simultaneously to another cell in the system if the distance between the two cells is larger than minimum reuse distance. When compared to the FCA, DCA has reduced the likelihood of blocking and even increased the trunking capacity of the network as all of the channels are available to all cells, i.e., good quality of service. But this type of assignment strategy results in heavy load on switching center at heavy traffic condition.

Handoff Process

When a user moves from one cell to the other, to keep the communication between the user pair, the user channel has to be shifted from one BS to the other without interrupting the call, i.e., when a MS moves into another cell, while the conversation is still in progress, the MSC automatically transfers the call to a new FDD channel without disturbing the conversation. This process is called as *handoff*. A schematic diagram of handoff is given in Figure 3.3.

Processing of handoff is an important task in any cellular system. Handoffs must

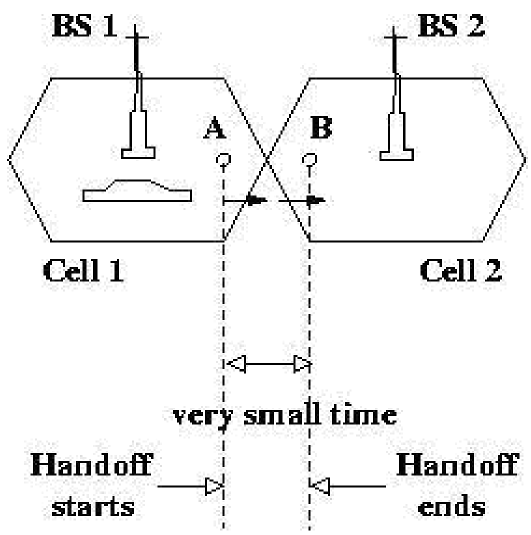


Figure 3.3: Handoff scenario at two adjacent cell boundary

Once a signal level is set as the minimum acceptable for good voice quality (Pr_{min}), then a slightly stronger level is chosen as the threshold (Pr_H) at which handoff has to be made, as shown in Figure 3.4. A parameter, called power margin, defined as

$$\Delta = Pr_H - Pr_{min} \tag{3.7}$$

is quite an important parameter during the handoff process since this margin Δ can neither be too large nor too small. If Δ is too small, then there may not be enough time to complete the handoff and the call might be lost even if the user crosses the cell boundary.

If Δ is too high o the other hand, then MSC has to be burdened with unnecessary handoffs. This is because MS may not intend to enter the other cell. Therefore Δ should be judiciously chosen to ensure imperceptible handoffs and to meet other objectives.

Factors Influencing Handoffs

The following factors influence the entire handoff process:

(a) Transmitted power: as we know that the transmission power is different for different cells, the handoff threshold or the power margin varies from cell to cell.

(b) Received power: depends on the Line of Sight (LoS) path between the user and the BS. Especially when the user the received power mostly is on the boundary of hexagonal cell.

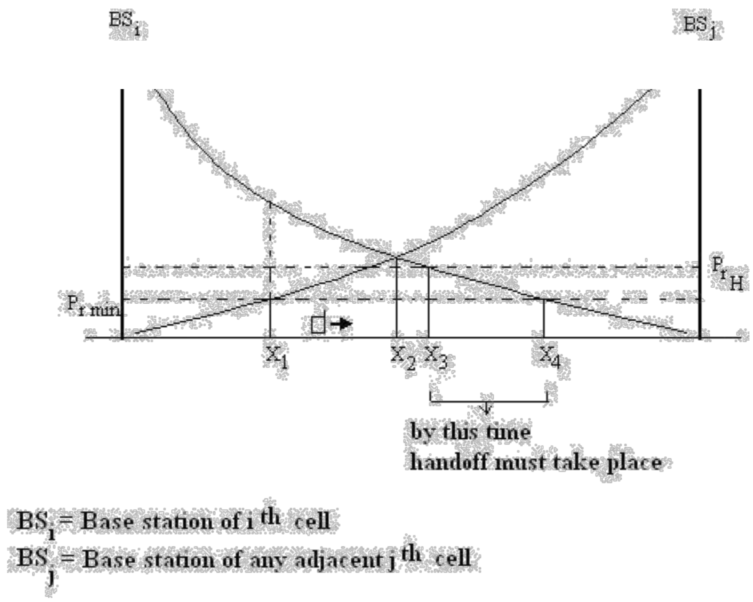


Figure 3.4: Handoff process associated with power levels, when the user is going from i-th cell to j-th cell.

the two cells, the LoS path plays a critical role in handoffs and therefore the power margin Δ depends on the minimum received power value from cell to cell.

(c) Area and shape of the cell: Apart from the power levels, the cell structure also a plays an important role in the handoff process.

(d) Mobility of users: The number of mobile users entering or going out of a particular cell, also fixes the handoff strategy of a cell.

To illustrate the reasons (c) and (d), let us consider a rectangular cell with sides $R1$ and $R2$ inclined at an angle θ with horizon, as shown in the Figure 3.5. Assume $N1$ users are having handoff in horizontal direction and $N2$ in vertical direction per unit length.

The number of crossings along $R1$ side is : $(N1 cos\theta + N2 sin\theta)R1$ and the number of crossings along $R2$ side is : $(N1 sin\theta + N2 cos\theta)R2$.

Then the handoff rate λH can be written as

$$\lambda H = (N1\cos\theta + N2\sin\theta)R1 + (N1\sin\theta + N2\cos\theta)R2. \quad (3.8)$$

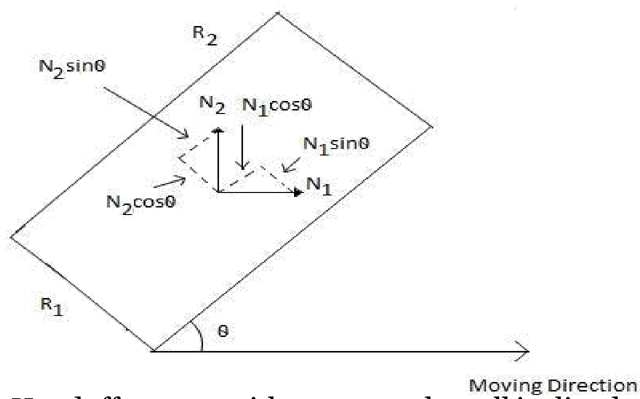


Figure 3.5: Handoff process with a rectangular cell inclined at an angle θ.

Now, given the fixed area $A = R_1R_2$, we need to find λ_H^{min} for a given θ. Replacing R_1 by $\frac{A}{R_2}$ and equating $\frac{d\lambda_H}{dR_1}$ to zero, we get

$$R_1^2 = A\left(\frac{N_1\sin\theta + N_2\cos\theta}{N_1\cos\theta + N_2\sin\theta}\right). \quad (3.9)$$

Similarly, for R_2, we get

$$R_2^2 = A\left(\frac{N_1\cos\theta + N_2\sin\theta}{N_1\sin\theta + N_2\cos\theta}\right). \quad (3.10)$$

From the above equations, we have $\lambda_H = 2\sqrt{A(N_1N_2 + (N_1^2 + N_2^2)\cos\theta\sin\theta)}$ which means it it minimized at $\theta = 0^o$. Hence $\lambda_H^{min} = 2\sqrt{AN_1N_2}$. Putting the value of θ in (3.9) or (3.10), we have $\frac{R_1}{R_2} = \frac{N_1}{N_2}$. This has two implications: (i) that handoff is minimized if rectangular cell is aligned with X-Y axis, i.e., $\theta = 0^o$, and, (ii) that the number of users crossing the cell boundary is inversely proportional to the dimension of the other side of the cell. The above analysis has been carried out for a simple square cell and it changes in more complicated way when we consider a hexagonal cell.

Handoffs In Different Generations

In 1G analog cellular systems, the signal strength measurements were made by the BS and in turn supervised by the MSC. The handoffs in this generation can be termed as Network Controlled Hand-Off (NCHO). The BS monitors the signal strengths of voice channels to determine the relative positions of the subscriber. The special receivers located on the BS are controlled by the MSC to monitor the signal strengths of the users in the neighboring cells which appear to be in need of handoff. Based on the information received from the special receivers the MSC decides whether a handoff is required or not. The approximate time needed to make a handoff successful was about 5-10 s. This requires the value of Δ to be in the order of 6dB to 12dB.

In the 2G systems, the MSC was relieved from the entire operation. In this generation, which started using the digital technology, handoff decisions were mobile assisted and therefore it is called Mobile Assisted Hand-Off (MAHO). In MAHO, the mobile center measures the power changes received from nearby base stations and notifies the two BS. Accordingly the two BS communicate and channel transfer occurs. As compared to 1G, the circuit complexity was increased here whereas the delay in handoff was reduced to 1-5 s. The value of Δ was in the order of 0-5 dB. However, even this amount of delay could create a communication pause.

In the current 3G systems, the MS measures the power from adjacent BS and automatically upgrades the channels to its nearer BS. Hence this can be termed as Mobile Controlled Hand-Off (MCHO). When compared to the other generations, delay during handoff is only 100 ms and the value of Δ is around 20 dBm. The Quality Of Service (QoS) has improved a lot although the complexity of the circuitry has further increased which is inevitable.

All these types of handoffs are usually termed as hard handoff as there is a shift in the channels involved. There is also another kind of handoff, called soft handoff, as discussed below.

Handoff in CDMA: In spread spectrum cellular systems, the mobiles share the same channels in every cell. The MSC evaluates the signal strengths received from different BS for a single user and then shifts the user from one BS to the other without actually changing the channel. These types of handoffs are called as soft handoff as there is no change in the channel.

Handoff Priority

While assigning channels using either FCA or DCA strategy, a guard channel concept must be followed to facilitate the handoffs. This means, a fraction of total available channels must be kept for handoff requests. But this would reduce the carried traffic and only fewer channels can be assigned for the residual users of a cell. A good solution to avoid such a dead-lock is to use DCA with handoff priority (demand based allocation).

A Few Practical Problems in Handoff Scenario

(a) Different speed of mobile users: with the increase of mobile users in urban areas, microcells are introduced in the cells to increase the capacity (this will be discussed later in this chapter). The users with high speed frequently crossing the micro-cells become burdened to MSC as it has to take care of handoffs. Several schemes thus have been designed to handle the simultaneous traffic of high speed and low speed users while minimizing the handoff intervention from the MSC, one of them being the 'Umbrella Cell' approach. This technique provides large area coverage to high speed users while providing small area coverage to users traveling at low speed. By using different antenna heights and different power levels, it is possible to provide larger and smaller cells at a same location. As illustrated in the Figure 3.6, umbrella cell is co-located with few other microcells. The BS can measure the speed of the user by its short term average signal strength over the RVC and decides which cell to handle that call. If the speed is less, then the corresponding microcell handles the call so that there is good corner coverage. This approach assures that handoffs are minimized for high speed users and provides additional microcell channels for pedestrian users.

(b) Cell dragging problem: this is another practical problem in the urban area with additional microcells. For example, consider there is a LOS path between the MS and BS1 while the user is in the cell covered by BS2. Since there is a LOS with the BS1, the signal strength received from BS1 would be greater than that received from BS2. However, since the user is in cell covered by BS2, handoff cannot take place and as a result, it experiences a lot of interferences. This problem can be solved by judiciously choosing the handoff threshold along with adjusting the coverage area.

(c) Inter-system handoff: if one user is leaving the coverage area of one MSC and is entering the area of another MSC, then the call might be lost if there is no handoff in this case too. Such a handoff is called inter-system handoff and in order to facilitate this, mobiles usually have roaming facility.

Interference & System Capacity

Susceptibility and interference problems associated with mobile communications equipment are because of the problem of time congestion within the electromagnetic spectrum. It is the limiting factor in the performance of cellular systems. This interference can occur from clash with another mobile in the same cell or because of a call in the adjacent cell. There can be interference between the base stations operating at same frequency band or any other non-cellular system's energy leaking inadvertently into the frequency band of the cellular system. If there is an interference in the voice channels, cross talk is heard will appear as noise between the users. The interference in the control channels leads to missed and error calls because of digital signaling. Interference is more severe in urban areas because of the greater RF noise and greater density of mobiles and base stations. The interference can be divided into 2 parts: co-channel interference and adjacent channel interference.

Co-channel interference (CCI)

For the efficient use of available spectrum, it is necessary to reuse frequency bandwidth over relatively small geographical areas. However, increasing frequency reuse also increases interference, which decreases system capacity and service quality. The cells where the same set of frequencies is used are call co-channel cells. Co-channel interference is the cross talk between two different radio transmitters using the same radio frequency as is the case with the co-channel cells. The reasons of CCI can be because of either adverse weather conditions or poor frequency planning or overly-crowded radio spectrum.

If the cell size and the power transmitted at the base stations are same then CCI will become independent of the transmitted power and will depend on radius of the cell (R) and the distance between the interfering co-channel cells (D). If D/R ratio is increased, then the effective distance between the co-channel cells will increase and interference will decrease. The parameter Q is called the frequency reuse ratio and is related to the cluster size. For hexagonal geometry

$$Q = \frac{D}{R} = \sqrt{3N} \qquad (3.11)$$

From the above equation, small of ‘Q’ means small value of cluster size ‘N’ and increase in cellular capacity. But large ‘Q’ leads to decrease in system capacity but increase in transmission quality. Choosing the options is very careful for the selection of ‘N’, the proof of which is given in the first section.

The Signal to Interference Ratio (SIR) for a mobile receiver which monitors the forward channel can be calculated as

$$\frac{S}{I} = \frac{S}{\sum_{i=1}^{i_0} I_i} \qquad (3.12)$$

where *i*0 is the number of co-channel interfering cells, S is the desired signal power from the baseband station and *Ii* is the interference power caused by the i-th interfer-ing co-channel base station. In order to solve this equation from power calculations, we need to look into the signal power characteristics. The average power in the mobile radio channel decays as a power law of the distance of separation between transmitter and receiver. The expression for the received power *Pr* at a distance d can be approximately calculated as

$$P_r = P_0 \left(\frac{d}{d_0} \right)^{-n} \qquad (3.13)$$

and in the dB expression as

$$P_r(\mathrm{dBm}) = P_0(\mathrm{dBm}) - 10n\log\left(\frac{d}{d_0} \right) \qquad (3.14)$$

$P0$ is the power received at a close-in reference point in the far field region at a small distance do from the transmitting antenna, and 'n' is the path loss exponent. Let us calculate the SIR for this system. If D_i is the distance of the i-th interferer from the mobile, the received power at a given mobile due to i-th interfering cell is proportional to $(D_i)^{-n}$ (the value of 'n' varies between 2 and 4 in urban cellular systems).

Let us take that the path loss exponent is same throughout the coverage area and the transmitted power be same, then SIR can be approximated as

$$\frac{S}{I} = \frac{R^{-n}}{\sum_{i=1}^{i_0} (D_i)^{-n}} \quad (3.15)$$

where the mobile is assumed to be located at R distance from the cell center. If we consider only the first layer of interfering cells and we assume that the interfer-ing base stations are equidistant from the reference base station and the distance between the cell centers is 'D' then the above equation can be converted as

$$\frac{S}{I} = \frac{(D/R)^n}{i_0} = \frac{(\sqrt{3N})^n}{i_0} \quad (3.16)$$

than or equal to 18 dB. If we take n=4 the value of 'N' can be calculated as 6.49. Therefore minimum N is 7. The above equations are based on hexagonal geometry and the distances from the closest interfering cells can vary if different frequency reuse plans are used.

We can go for a more approximate calculation for co-channel SIR. This is the example of a 7 cell reuse case. The mobile is at a distance of D-R from 2 closest interfering cells and approximately D+R/2, D, D-R/2 and D+R distance from other interfering cells in the first tier. Taking n = 4 in the above equation, SIR can be approximately calculated as which is an approximate measure of the SIR. Subjective tests performed on AMPS cellular system which uses FM and 30 kHz channels show that sufficient voice quality can be obtained by SIR being greater

$$\frac{S}{I} = \frac{R^{-4}}{2(D-R)^{-4} + (D+R)^{-4} + (D)^{-4} + (D+R/2)^{-4} + (D-R/2)^{-4}} \qquad (3.17)$$

which can be rewritten in terms frequency reuse ratio Q as

$$\frac{S}{I} = \frac{1}{2(Q-1)^{-4} + (Q+1)^{-4} + (Q)^{-4} + (Q+1/2)^{-4} + (Q-1/2)^{-4}}. \qquad (3.18)$$

Using the value of N equal to 7 (this means Q = 4.6), the above expression yields that worst case SIR is 53.70 (17.3 dB). This shows that for a 7 cell reuse case the worst case SIR is slightly less than 18 dB. The worst case is when the mobile is at the corner of the cell i.e., on a vertex as shown in the Figure 3.6. Therefore N = 12 cluster size should be used. But this reduces the capacity by 7/12 times. Therefore, co-channel interference controls link performance, which in a way controls frequency reuse plan and the overall capacity of the cellular system. The effect of co-channel interference can be minimized by optimizing the frequency assignments of the base stations and their transmit powers. Tilting the base-station antenna to limit the spread of the signals in the system can also be done.

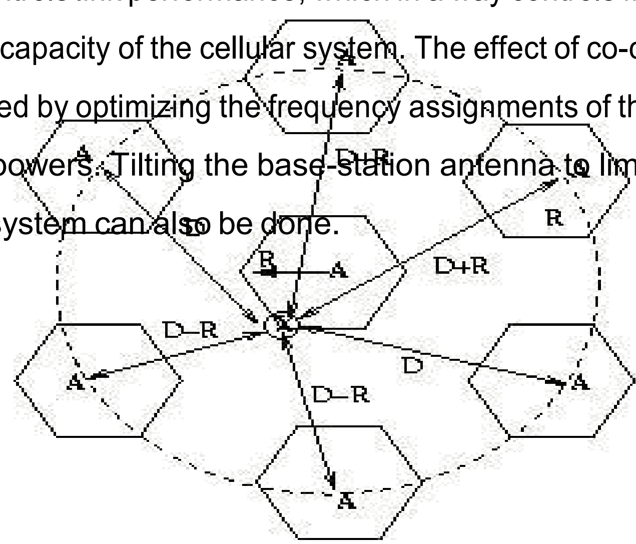


Figure 3.6: First tier of co-channel interfering cells

Adjacent Channel Interference (ACI)

This is a different type of interference which is caused by adjacent channels i.e. channels in adjacent cells. It is the signal impairment which occurs to one frequency due to presence of another signal on a nearby frequency. This occurs when imperfect receiver filters allow nearby frequencies to leak into the pass band. This problem is enhanced if the adjacent channel user is transmitting in a close range compared to the subscriber's receiver while the receiver attempts to receive a base station on the channel. This is called near-far effect. The more adjacent channels are packed into the channel block, the higher the spectral efficiency, provided that the performance degradation can be tolerated in the system link budget. This effect can also occur if a mobile close to a base station transmits on a channel close to one being used by a weak mobile. This problem might occur if the base station has problem in discriminating the mobile user from the "bleed over" caused by the close adjacent channel mobile.

Adjacent channel interference occurs more frequently in small cell clusters and heavily used cells. If the frequency separation between the channels is kept large this interference can be reduced to some extent. Thus assignment of channels is given such that they do not form a contiguous band of frequencies within a particular cell and frequency separation is maximized. Efficient assignment strategies are very much important in making the interference as less as possible. If the frequency fac-tor is small then distance between the adjacent channels cannot put the interference level within tolerance limits. If a mobile is 10 times close to the base station than other mobile and has energy spill out of its pass band, then SIR for weak mobile is approximately.

$$\frac{S}{I} = 10^{-n} \qquad (3.19)$$

which can be easily found from the earlier SIR expressions. If n = 4, then SIR is −52 dB. Perfect base station filters are needed when close-in and distant users share the same cell. Practically, each base station receiver is preceded by a high Q cavity filter in order to remove adjacent channel interference. Power control is also very much important for the prolonging of the battery life for the subscriber unit but also reduces reverse channel SIR in the system. Power control is done such that each mobile transmits the lowest power required to maintain a good quality link on the reverse channel.

Power Conrol For Reducing Interference

In practical cellular radio and personal communication systems the power levels transmitted by every subscriber unit are under constant control by the serving base stations. This is done to ensure that each mobile transmits the smallest power necessary to maintain a good quality link on the reverse channel. Power control not only helps prolong battery life for subscriber unit,but also dramatically reduces the reverse channel S/I in the system.

Trunking and Grade of Service

In the previous sections, we have discussed the frequency reuse plan, the design trade-offs and also explored certain capacity expansion techniques like cell-splitting and sectoring. Now, we look at the relation between the number of radio channels a cell contains and the number of users a cell can support. Cellular systems use the concept of trunking to accommodate a large number of users in a limited radio spectrum. It was found that a central office associated with say, 10,000 telephones requires about 50 million connections to connect every possible pair of users. However, a worst case maximum of 5000 connections need to be made among these telephones at any given instant of time, as against the possible 50 million connections. In fact, only a few hundreds of lines are needed owing to the relatively short duration of a call. This indicates that the resources are shared so that the number of lines is much smaller than the number of possible connections. A line that connects switching offices and that is shared among users on an as-needed basis is called a trunk.

The fact that the number of trunks needed to make connections between offices is much smaller than the maximum number that could be used suggests that at times there might not be sufficient facilities to allow a call to be completed. A call that cannot be completed owing to a lack of resources is said to be blocked. So one important to be answered in mobile cellular systems is: How many channels per cell are needed in a cellular telephone system to ensure a reasonably low probability that a call will be blocked?

In a trunked radio system, a channel is allotted on per call basis. The performance of a radio system can be estimated in a way by looking at how efficiently the calls are getting connected and also how they are being maintained at handoffs.

Some of the important factors to take into consideration are (i) Arrival statistics, (ii)Service statistics, (iii)Number of servers/channels.

Let us now consider the following assumptions for a bufferless system handling 'L' users as shown in Figure 3.11:

(i) The number of users L is large when compared to 1.

(ii) Arrival statistics is Poisson distributed with a mean parameter λ.

(iii) Duration of a call is exponentially distributed with a mean rate $\mu 1$.

(iv) Residence time of each user is exponentially distributed with a rate parameter $\mu 2$

(v) The channel holding rate therefore is exponentially distributed with a parameter

$\mu = \mu 1 + \mu 2$.

(vi) There is a total of 'J' number of channels ($J \leq L$).

To analyze such a system, let us recapitulate a queuing system in brief. Consider an M/M/m/m system which is an m-server loss system. The name M/M/m/m reflects

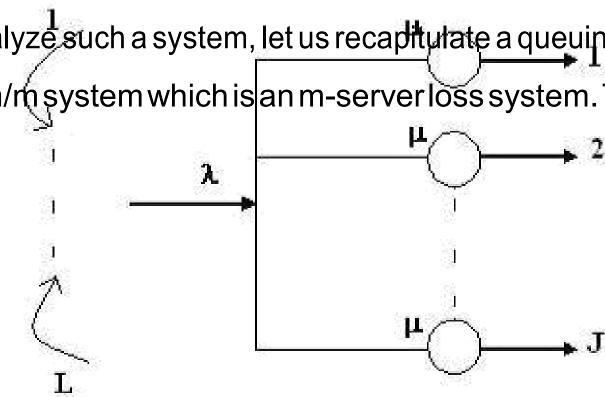


Figure 3.11: The bufferless J-channel trunked radio system

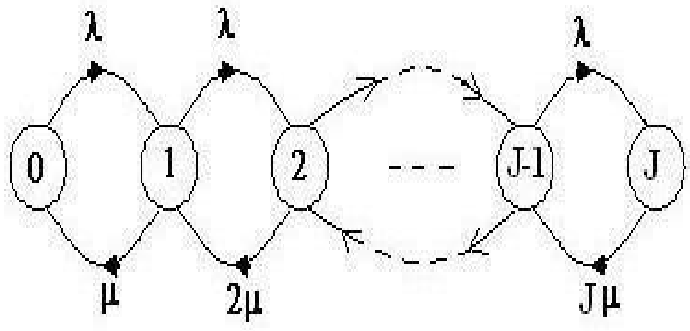
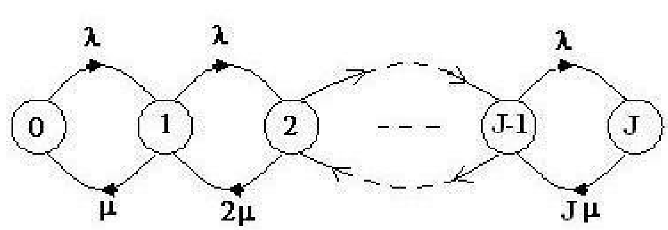


Figure 3.12:Discrete-time Markov chain for the M/M/J/J trunked radio system.



standard queuing theory nomenclature whereby:

(i) the first letter indicates the nature of arrival process(e.g. M stands for memory-less which here means a Poisson process).
(ii) the second letter indicates the nature of probability distribution of service times. (e.g M stands for exponential distribution). In all cases,successive inter ar-rival times and service times are assumed to be statistically independent of each other.
(iii) the third letter indicates the number of servers.
(iv) the last letter indicates that if an arrival finds all 'm' users to be busy, then it will not enter the system and is lost.

In view of the above, the bufferless system as shown in Figure 3.11 can be modeled as M/M/J/J system and the discrete-time Markov chain of this system is shown in Figure 3.12.

Trunking mainly exploits the statistical behavior of users so that a fixed number of channels can be used to accommodate a large, random user community. As the number of telephone lines decrease, it becomes more likely that all channels are busy for a particular user. As a result, the call gets rejected and in some systems, a queue may be used to hold the caller's request until a channel becomes available. In the telephone system context the term Grade of Service (GoS) is used to mean the probability that a user's request for service will be blocked because a required facility, such as a trunk or a cellular channel, is not available. For example, a GoS of 2 % implies that on the average a user might not be successful in placing a call on 2 out of every 100 attempts. In practice the blocking frequency varies with time. One would expect far more call attempts during business hours than during the middle of the night. Telephone operating companies maintain usage records and can identify a "busy hour", that is, the hour of the day during which there is the greatest demand for service. Typically, telephone systems are engineered to provide a specified grade of service during a specified busy hour.

User calling can be modeled statistically by two parameters: the average number of call requests per unit time λ_{user} and the average holding time H. The parameter λ_{user} is also called the average arrival rate, referring to the rate at which calls from a single user arrive. The average holding time is the average duration of a call. The product:

$$A_{user} = \lambda_{user} H \tag{3.26}$$

that is, the product of the average arrival rate and the average holding time–is called the offered traffic intensity or offered load. This quantity represents the average traffic that a user provides to the system. Offered traffic intensity is a quantity that is traditionally measured in Erlangs. One Erlang represents the amount of traffic intensity carried by a channel that is completely occupied. For example, a channel that is occupied for thirty minutes during an hour carries 0.5 Erlang of traffic.

Call arrivals or requests for service are modeled as a Poisson random process. It is based on the assumption that there is a large pool of users who do not cooperate in deciding when to place calls. Holding times are very well predicted using an exponential probability distribution. This implies that calls of long duration are much less frequent than short calls. If the traffic intensity offered by a single user is A_{user}, then the traffic intensity offered by N users is $A = NA_{user}$. The purpose of the statistical model is to relate the offered traffic intensity A, the grade of service P_b, and the number of channels or trunks C needed to maintain the desired grade of service.

Two models are widely used in traffic engineering to represent what happens when a call is blocked. The blocked calls cleared model assumes that when a channel or trunk is not available to service an arriving call, the call is cleared from the system. The second model is known as blocked calls delayed. In this model a call that cannot be serviced is placed on a queue and will be serviced when a channel or trunk becomes available.

Use of the blocked-calls-cleared statistical model leads to the Erlang B formula that relates offered traffic intensity A, grade of service P_b, and number of channels K. The Erlang B formula is:

$$P_b = \frac{A^K/K!}{\sum_{n=0}^{K} A^n/n!} \qquad (3.27)$$

When the blocked-calls-delayed model is used, the "grade of service" refers to the probability that a call will be delayed. In this case the statistical model leads to the Erlang C formula,

$$P[delay] = \frac{A^K/[(K-A)(K-1)]!}{A^K/[(K-A)(K-1)]! + \sum_{n=0}^{K} A^n/n!} \qquad (3.28)$$

Improving Coverage and Capacity in Cellular Systems

Previously, we have seen that the frequency reuse technique in cellular systems allows for almost boundless expansion of geographical area and the number of mobile system users who could be accommodated. In designing a cellular layout, the two parameters which are of great significance are the cell radius R and the cluster size N, and we have also seen that co-channel cell distance $D = \sqrt{3N}R$. In the following, a brief description of the design trade-off is given, in which the above two parameters play a crucial role.

The cell radius governs both the geographical area covered by a cell and also the number of subscribers who can be serviced, given the subscriber density. It is easy to see that the cell radius must be as large as possible. This is because, every cell requires an investment in a tower, land on which the tower is placed, and radio transmission equipment and so a large cell size minimizes the cost per subscriber. Eventually, the cell radius is determined by the requirement that adequate signal to noise ratio be maintained over the coverage area. The SNR is determined by several factors such as the antenna height, transmitter power, receiver noise figure etc. Given a cell radius R and a cluster size *N*, the geographic area covered by a cluster is

$$A_{cluster} = NA_{cell} = N3\sqrt{3}R^2/2. \qquad (3.20)$$

If the total serviced area is A_{total}, then the number of clusters M that could be accommodated is given by

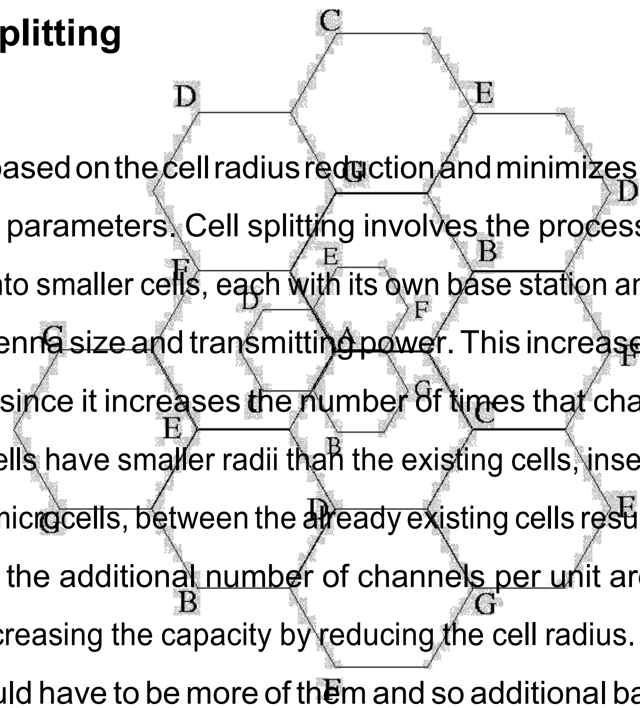
$$M = A_{total}/A_{cluster} = A_{total}/(N3\sqrt{3}R^2/2). \qquad (3.21)$$

Note that all of the available channels N, are reused in every cluster. Hence, to make the maximum number of channels available to subscribers, the number of clusters M should be large, which, by Equation (3.21), shows that the cell radius should be small. However, cell radius is determined by a trade-off: R should be as large as possible to minimize the cost of the installation per subscriber, but R should be as small as possible to maximize the number of customers that the system can accommodate. Now, if the cell radius R is fixed, then the number of clusters could be maximized by minimizing the size of a cluster *N*. We have seen earlier that the size of a cluster depends on the frequency reuse ratio Q. Hence, in determining the value of *N*, another trade-off is encountered in that *N* must be small to accommodate large number of subscribers, but should be sufficiently large so as to minimize the interference effects.

Now, we focus on the issues regarding system expansion. The history of cellular phones has been characterized by a rapid growth and expansion in cell subscribers. Though a cellular system can be expanded by simply adding cells to the geographical area, the way in which user density can be increased is also important to look at. This is because it is not always possible to counter the increasing demand for cellular systems just by increasing the geographical coverage area due to the limitations in obtaining new land with suitable requirements. We discuss here two methods for dealing with an increasing subscriber density: Cell Splitting and Sectoring. The other method, microcell zone concept can treated as enhancing the QoS in a cellular system.

The basic idea of adopting the cellular approach is to allow space for the growth of mobile users. When a new system is deployed, the demand for it is fairly low and users are assumed to be uniformly distributed over the service area. However, as new users subscribe to the cellular service, the demand for channels may begin to exceed the capacity of some base stations. As discussed previously, the number of channels available to customers (equivalently, the channel density per square kilometer) could be increased by decreasing the cluster size. However, once a system has been initially deployed, a system-wide reduction in cluster size may not be necessary since user density does not grow uniformly in all parts of the geographical area. It might be that an increase in channel density is required only in specific parts of the system to support an increased demand in those areas. Cell-splitting is a technique which has the capability to add new smaller cells in specific areas of the system.

Cell-Splitting



Cell Splitting is based on the cell radius reduction and minimizes the need to modify the existing cell parameters. Cell splitting involves the process of sub-dividing a congested cell into smaller cells, each with its own base station and a corresponding reduction in antenna size and transmitting power. This increases the capacity of a cellular system since it increases the number of times that channels are reused. Since the new cells have smaller radii than the existing cells, inserting these smaller cells, known as microcells, between the already existing cells results in an increase of capacity due to the additional number of channels per unit area. There are few challenges in increasing the capacity by reducing the cell radius. Clearly, if cells are small, there would have to be more of them and so additional base stations will be needed in the system. The challenge in this case is to introduce the new base stations without the need to move the already existing base station towers. The other challenge is to meet the generally increasing demand that may vary quite rapidly between geographical areas of the system. For instance, a city may have highly populated areas and so the demand must be supported by cells with the smallest radius. The radius of cells will generally increase as we move from urban to sub urban areas, because the user density decreases on moving towards sub-urban areas. The key factor is to add as minimum number of smaller cells as possible wherever an increase in demand occurs. The gradual addition of the smaller cells implies that, at least for a time, the cellular system operates with cells of more than one size.

Figure 3.7: Splitting of congested seven-cell clusters.

Figure 3.7 shows a cellular layout with seven-cell clusters. Consider that the cells in the center of the diagram are becoming congested, and cell A in the center has reached its maximum capacity. Figure also shows how the smaller cells are being superimposed on the original layout. The new smaller cells have half the cell radius of the original cells. At half the radius, the new cells will have one-fourth of the area and will consequently need to support one-fourth the number of subscribers. Notice that one of the new smaller cells lies in the center of each of the larger cells. If we assume that base stations are located in the cell centers, this allows the original base stations to be maintained even in the new system layout. However, new base stations will have to be added for new cells that do not lie in the center of the larger cells. The organization of cells into clusters is independent of the cell radius, so that the cluster size can be the same in the small-cell layout as it was in the large-cell layout. Also the signal-to-interference ratio is determined by cluster size and not by cell radius. Consequently, if the cluster size is maintained, the signal-to-interference ratio will be the same after cell splitting as it was before. If the entire system is replaced with new half-radius cells, and the cluster size is maintained, the number of channels per cell will be exactly as it was before, and the number of subscribers per cell ill have been reduced.

When the cell radius is reduced by a factor, it is also desirable to reduce the transmitted power. The transmit power of the new cells with radius half that of the old cells can be found by examining the received power PR at the new and old cell boundaries and setting them equal. This is necessary to maintain the same frequency re-use plan in the new cell layout as well. Assume that PT1 and PT2 are the transmit powers of the larger and smaller base stations respectively. Then, assuming a path loss index n=4, we have power received at old cell boundary = $PT1/R^4$ and the power received at new cell boundary = $PT2/(R/2)^4$. On equating the two received powers, we get $PT2 = PT1 / 16$. In other words, the transmit power must be reduced by 12 dB in order to maintain the same S/I with the new system lay-out.

At the beginning of this channel splitting process, there would be fewer channels in the smaller power groups. As the demand increases, more and more channels need to be accommodated and hence the splitting process continues until all the larger cells have been replaced by the smaller cells, at which point splitting is complete within the region and the entire system is rescaled to have a smaller radius per cell. If a cellular layout is replaced entirety by a new layout with a smaller cell radius, the signal-to-interference ratio will not change, provided the cluster size does not change. Some special care must be taken, however, to avoid co-channel interference when both large and small cell radii coexist. It turns out that the only way to avoid interference between the large-cell and small-cell systems is to assign entirely different sets of channels to the two systems. So, when two sizes of cells co-exist in a system, channels in the old cell must be broken down into two groups, one that corresponds to larger cell reuse requirements and the other which corresponds to the smaller cell reuse requirements. The larger cell is usually dedicated to high speed users as in the umbrella cell approach so as to minimize the number of hand-offs.

Sectoring

Sectoring is basically a technique which can increase the SIR without necessitating an increase in the cluster size. Till now, it has been assumed that the base station is located in the center of a cell and radiates uniformly in all the directions behaving as an omni-directional antenna. However it has been found that the co-channel interference in a cellular system may be decreased by replacing a single omni-directional antenna at the base station by several directional antennas, each radiating within a specified sector. In the Figure 3.8, a cell is shown which has been split into three 120^o sectors. The base station feeds three 120^o directional antennas, each of which radiates into one of the three sectors. The channel set serving this cell has also been divided, so that each sector is assigned one-third of the available number cell of channels. This technique for reducing co-channel interference wherein by using suitable

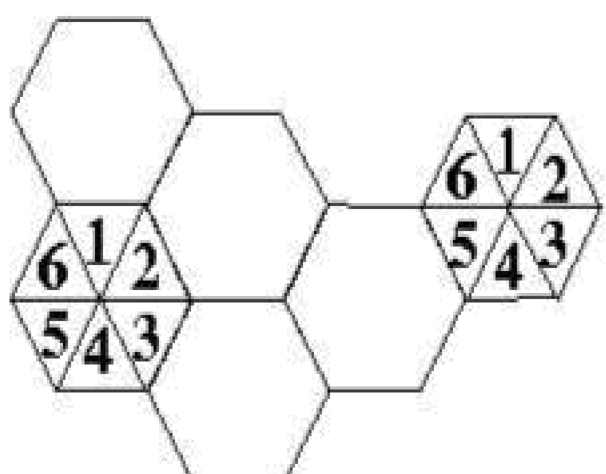


Figure 3.9: A seven-cell cluster with 60^o sectors.

directional antennas, a given cell would receive interference and transmit with a fraction of available co-channel cells is called ’sectoring’. In a seven-cell-cluster layout with 120^o sectored cells, it can be easily understood that the mobile units in a particular sector of the center cell will receive co-channel interference from only two of the first-tier co-channel base stations, rather than from all six. Likewise, the base station in the center cell will receive co-channel interference from mobile units in only two of the co-channel cells. Hence the signal to interference ratio is now modified to

$$\frac{S}{I} = \frac{(\sqrt{3N})^n}{2} \qquad (3.22)$$

where the denominator has been reduced from 6 to 2 to account for the reduced number of interfering sources. Now, the signal to interference ratio for a seven-cell cluster layout using 120^o sectored antennas can be found from equation (3.24) to be 23.4 dB which is a significant improvement over the Omni-directional case where the worst-case S/I is found to be 17 dB (assuming a path-loss exponent, n=4). Some cellular systems divide the cells into 60^o sectors. Similar analysis can be performed on them as well.

Microcell Zone Concept

The increased number of handoffs required when sectoring is employed results in an increased load on the switching and control link elements of the mobile system. To overcome this problem, a new microcell zone concept has been proposed. As shown in Figure 3.10, this scheme has a cell divided into three microcell zones, with each of the three zone sites connected to the base station and sharing the same radio equipment. It is necessary to note that all the microcell zones, within a cell, use the same frequency used by that cell; that is no handovers occur between microcells. Thus when a mobile user moves between two microcell zones of the cell, the BS simply switches the channel to a different zone site and no physical re-allotment of channel takes place.

Locating the mobile unit within the cell: An active mobile unit sends a signal to all zone sites, which in turn send a signal to the BS. A zone selector at the BS uses that signal to select a suitable zone to serve the mobile unit - choosing the zone with the strongest signal.

Base Station Signals: When a call is made to a cellular phone, the system already knows the cell location of that phone. The base station of that cell knows in which zone, within that cell, the cellular phone is located. Therefore when it receives the signal, the base station transmits it to the suitable zone site. The zone site receives the cellular signal from the base station and transmits that signal to the mobile phone after amplification. By confining the power transmitted to the mobile phone, co-channel interference is reduced between the zones and the capacity of system is increased.

Benefits of the micro-cell zone concept: 1) Interference is reduced in this case as compared to the scheme in which the cell size is reduced.

2) Handoffs are reduced (also compared to decreasing the cell size) since the micro-cells within the cell operate at the same frequency; no handover occurs when the mobile unit moves between the microcells.

3) Size of the zone apparatus is small. The zone site equipment being small can be mounted on the side of a building or on poles.

4) System capacity is increased. The new microcell knows where to locate the mobile unit in a particular zone of the cell and deliver the power to that zone. Since

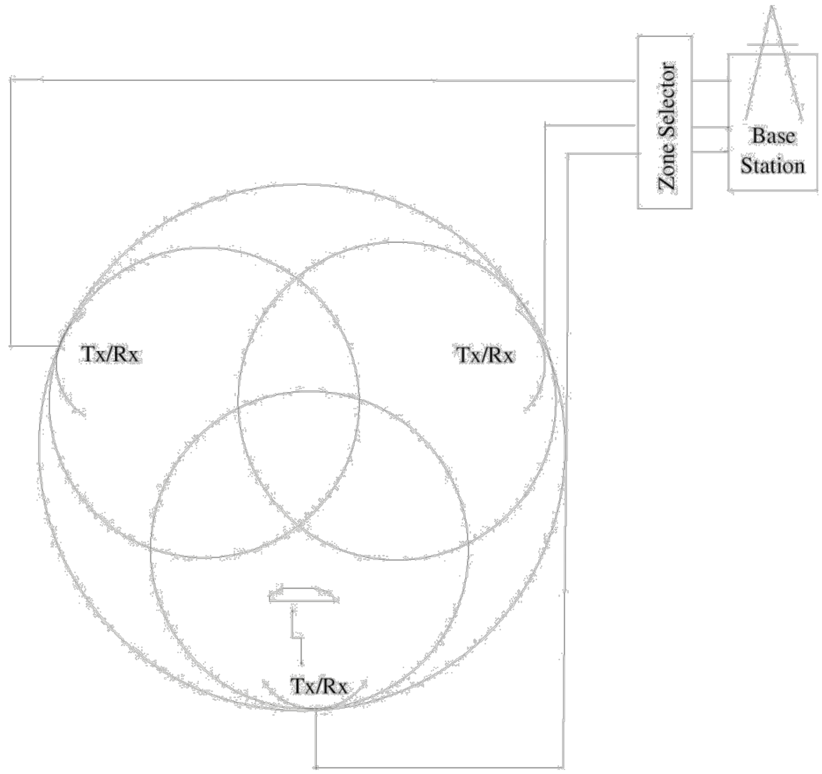


Figure 3.10: The micro-cell zone concept

the signal power is reduced, the microcells can be closer and result in an increased system capacity. However, in a microcellular system, the transmitted power to a mobile phone within a microcell has to be precise; too much power results in interference between microcells, while with too little power the signal might not reach the mobile phone. This is a drawback of microcellular systems, since a change in the surrounding (a new building, say, within a microcell) will require a change of the transmission power.

Unit -2 Mobile Radio Propagation:

Large-Scale Path Loss

Introduction

There are two basic ways of transmitting an electro-magnetic (EM) signal, through a guided medium or through an unguided medium. Guided mediums such as coaxial cables and fiber optic cables, are far less hostile toward the information carrying EM signal than the wireless or the unguided medium. It presents challenges and conditions which are unique for this kind of transmissions. A signal, as it travels through the wireless channel, undergoes many kinds of propagation effects such as reflection, diffraction and scattering, due to the presence of buildings, mountains and other such obstructions. Reflection occurs when the EM waves impinge on objects which are much greater than the wavelength of the traveling wave. Diffraction is a phenomena occurring when the wave interacts with a surface having sharp irregularities. Scattering occurs when the medium through the wave is traveling contains objects which are much smaller than the wavelength of the EM wave. These varied phenomena's lead to large scale and small scale propagation losses. Due to the inherent randomness associated with such channels they are best described with the help of statistical models. Models which predict the mean signal strength for arbitrary transmitter receiver distances are termed as large scale propagation models. These are termed so because they predict the average signal strength for large Tx-Rx separations, typically for hundreds of kilometers.

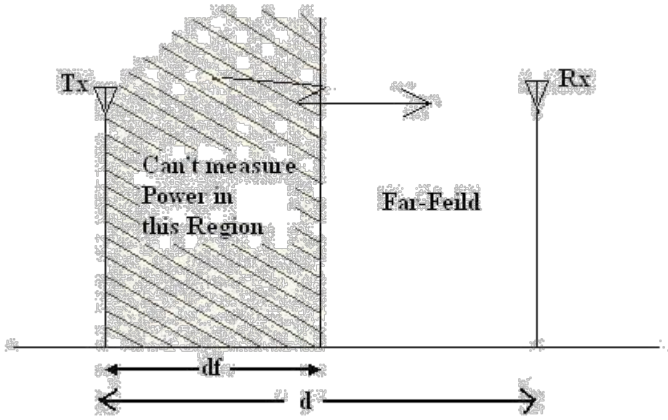


Figure 4.1: Free space propagation model, showing the near and far fields.

Free Space Propagation Model

Although EM signals when traveling through wireless channels experience fading effects due to various effects, but in some cases the transmission is with a direct line of sight such as in satellite communication. Free space model predicts that the received power decays as negative square root of the distance. Friis free space equation is given by

$$P_r(d) = \frac{P_t tt_t tt_r \lambda^2}{(4\pi)^2 d^2 L} \tag{4.1}$$

where Pt is the transmitted power, $Pr(d)$ is the received power, ttt is the transmitter antenna gain, ttr is the receiver antenna gain, d is the Tx-Rx separation and L is the system loss factor depended upon line attenuation, filter losses and antenna losses and not related to propagation. The gain of the antenna is related to the effective aperture of the antenna which in turn is dependent upon the physical size of the antenna as given below

$$tt = 4\pi Ae/\lambda^2. \tag{4.2}$$

The path loss, representing the attenuation suffered by the signal as it travels through the wireless channel is given by the difference of the transmitted and received power in dB and is expressed as:

$$PL(dB) = 10 \log Pt/Pr. \tag{4.3}$$

The fields of an antenna can broadly be classified in two regions, the far field and the near field. It is in the far field that the propagating waves act as plane waves and the power decays inversely with distance. The far field region is also termed as Fraunhofer region and the Friis equation holds in this region. Hence, the Friis equation is used only beyond the far field distance, d_f, which is dependent upon the largest dimension of the antenna as

$$d_f = 2D^2/\lambda. \quad (4.4)$$

Also we can see that the Friis equation is not defined for d=0. For this reason, we use a close in distance, d_o, as a reference point. The power received, $P_r(d)$, is then given by:

$$P_r(d) = P_r(d_o)(d_o/d)^2. \quad (4.5)$$

Ex. 1: Find the far field distance for a circular antenna with maximum dimension of 1 m and operating frequency of 900 MHz.

Solution: Since the operating frequency f = 900 Mhz, the wavelength

$$\lambda = \frac{3 \times 10^8 m/s}{900 \times 10^6 Hz} m$$

. Thus, with the largest dimension of the antenna, D=1m, the far field distance is

$$d_f = \frac{2D^2}{\lambda} = \frac{2(1)^2}{0.33} = 6m$$

.

Ex. 2: A unit gain antenna with a maximum dimension of 1 m produces 50 W power at 900 MHz. Find (i) the transmit power in dBm and dB, (ii) the received power at a free space distance of 5 m and 100 m. Solution:

(i) Tx power = 10log(50) = 17 dB = (17+30) dBm = 47 dBm

(ii) $d_f = \frac{2 \times D^2}{\lambda} = \frac{2 \times 1^2}{1/3} = 6m$

Thus the received power at 5 m can not be calculated using free space distance formula.

At 100 m ,

$$P_R = \frac{P_T G_T G_R \lambda^2}{4\pi d^2}$$

$$= \frac{50 \times 1 \times (1/3)^2}{4\pi 100^2}$$

Z.S Relating Power to Electrlc Field

Z.S Relating Power to Electrlc Field

The free space path loaa model of Section 3.2 is readily derived from first

a.s Rdistiñg Porter to nlectrlc **Held**

The free space path losa model of Section fl.Z is readily derived from firet principles. It can be proven that any radiating structure produces electric and magnetic fields f Om87 J, fKra5OJ. Consider a small linear radiator of length L , than ia placed cnincident with the z-axis and has its center at tire origin, as shown in Figure 8.2.

3.3 Relating Power to Electric Field

The free space path IOB8 model of Section 3.2 ia readily derived from first principles. It can be proven that any radiating structure produced electric and magnetic fields [Gn87), fKra50l Consider s amall linear radiator of length L , that ia placed coincident with the z-axis and has its center at the origin, as shown in Figure 3.2.

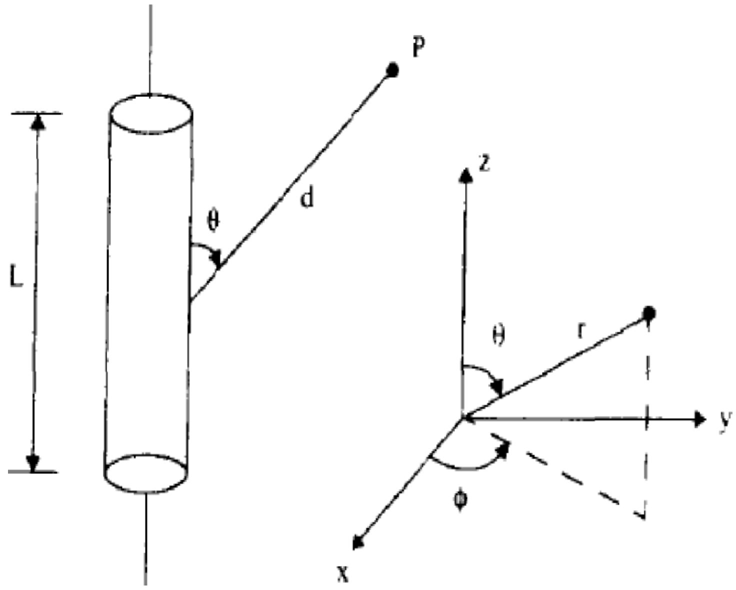


Figure 3.2
1]l uatr fiion of a linear radiator of length L (L ‹• 8,1 carrying a current of amplitude ip and asking an angle 8 with a point. at distance d.

If a current flows through such an antenna, it launches electric and magnetic fields that can be expressed as

$$E_r \quad \frac{i_0 L\cos\theta}{2\pi\varepsilon_0 c}\left\{\frac{1}{d^2} \quad \frac{}{j\omega_c d^3}\right\} e^{\text{yr› i i d c i}} \qquad \text{f3.10)}$$

$$E_\theta \quad \frac{i_0 L\sin\theta}{4\pi\varepsilon_0 c^2}\left\{\frac{j\omega_c}{d} + \frac{c}{d^2} + \frac{c^2}{j\omega_c d^3}\right\} e^{-j\omega_c(t \cdot d \; c)} \qquad (3.11)$$

$$H_\phi \quad \frac{i_0 L\sin\theta}{4\pi c}\left\{\frac{j\omega_c}{d} + \frac{c}{d^2}\right\} e^{j\omega_c(t \;\; d \;\; c)} \qquad (3.12)$$

with Ed = H -- H g -- $ñ$. In the above equations. all i d terms represent the radiation field component, all I d-' terms represent the induction field component, and all I "d°' terms represent the electrostatic field component. As seen from equations I:3.10a to 13.12). the electrostatic end inductive fields decay much faster with distance than the radiation field. At regions far **away** from the **transmitter** lfar-field region) the electrostatic a nd inductive fields become negligible and only the radiated field components of *Eg* and H need be considered.

In free space, the *power flue density Pq* lexpressed in W/m ^{2}) is given by

$$P_d \quad \frac{EIRP}{4\pi d^2} = \frac{P_t G_t}{4\pi d^2} \quad \frac{E^2}{R_{fs}} \quad \frac{E^{\cdot}}{\eta}\ \text{W/m}^2 \qquad (3.\ \text{T3›}$$

w'here fi¿, is the intrinsic impedance of free space given by q = l 20a n isr7ni. Thus, the power dux density is

$$P_d \quad \frac{\text{Sri'}}{377\Omega} \text{ W/m}^2 \qquad (3.14)$$

where .f'| represents the magnit adc of the radiating portion of the electric field in the for field. Pigure 8.pa illustrates how the power flux density disperses in free space from an isotropic point source. *Pg* may be thought Of as the *EIRP* divided by the surface area of a sphere with radius *d*. The power received at distance *d*, "r '"'• is given by the power Oux density times the effective aperture o1 the receiver antenna, and can be related to the electric field using equations (8,1), (S. 2), (3.13), and f3.14).

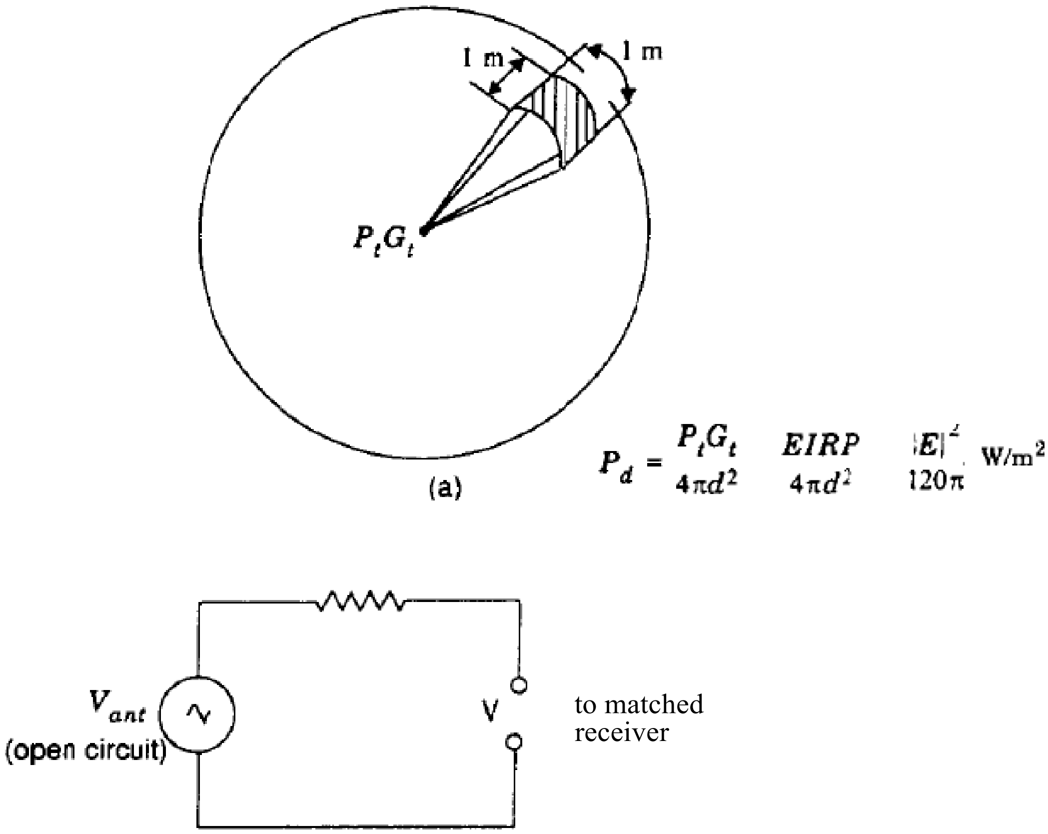


Figure 3.3
(a) Power flux denBity cl a distance *d* frum a point aource.
fb) Mndel for voltage applied to the input. of a mccc ver.

$$P \;\langle d \quad - P_d A \quad - \frac{|E|^2}{120\pi} A_e \;'' \quad \frac{P_tG_tG_r\lambda^2}{(4\pi)^2 d^2} \text{ Watts} \qquad (315)$$

Equation (3.15) relates electric field (w4th units of V/m) to received power (with units of watts , and i» identical to equation (3.11 Cth L 1

Often it is use5il tn relate the received power level to s receiver input shortage, as well as to an induced E-field at the retei ver antenna. If the receiver antenna is modeled as a matched resistive load to the receiver, then the receive antenna will induce an rms voltage into the receiver which is half of the open cir cuit voltage at the antenna. Thus, if V is the rms voltage at the input of a receiver (measured by a high impedance voltmeter), and R_{ant} is the resistance o the matched receiver, the received power is given by

$$P_r(d) = \frac{V^2}{R_{ant}} = \frac{[V_{ant}/2]^2}{R_{ant}} = \frac{V_{ant}^2}{4R_{ant}} \tag{3.16}$$

Through equations (3.14) to (3.16), it is possible to relate the received powe to the received E-field or the open circuit rms voltage at the receiver antenna ter minals. Figure 3.3b illustrates an equivalent circuit model. Note $V_{ant} = V$ wher there is no load.

Basic Methods of Propagation

Reflection, diffraction and scattering are the three fundamental phenomena that cause signal propagation in a mobile communication system, apart from LoS com-munication. The most important parameter, predicted by propagation models based on above three phenomena, is the received power. The physics of the above phe-nomena may also be used to describe small scale fading and multipath propagation. The following subsections give an outline of these phenomena.

Reflection

Reflection occurs when an electromagnetic wave falls on an object, which has very large dimensions as compared to the wavelength of the propagating wave. For ex-ample, such objects can be the earth, buildings and walls. When a radio wave falls on another medium having different electrical properties, a part of it is transmitted into it, while some energy is reflected back. Let us see some special cases. If the medium on which the e.m. wave is incident is a dielectric, some energy is reflected back and some energy is transmitted. If the medium is a perfect conductor, all energy is reflected back to the first medium. The amount of energy that is reflected back depends on the polarization of the e.m. wave.

Another particular case of interest arises in parallel polarization, when no re-flection occurs in the medium of origin. This would occur, when the incident angle would be such that the reflection coefficient is equal to zero. This angle is the Brewster's angle. By applying laws of electro-magnetics, it is found to be

$$\sin(\theta_B) = \sqrt{\frac{s_1}{s_1 + s_2}} . \tag{4.6}$$

Further, considering perfect conductors, the electric field inside the conductor is always zero. Hence all energy is reflected back. Boundary conditions require that

$$\theta_i = \theta_r \tag{4.7}$$

and

$$E_i = E_r \tag{4.8}$$

for vertical polarization, and

$$E_i = -E_r \tag{4.9}$$

for horizontal polarization.

Reflection from Dielectrics

Figure 3.4 showfl as electromagnetic wave incident at an anple B, Frith the plane of the boundary between two dielectric media. As shpwn in the figure, part of the energy is reflected hack to the first media at en angle I, , and part of the energy is transmitted frefractedl into the aecond media at an angle 8, . The nature of reflection varies with the direction of polarization of the E-field. The behavior for arbitrary directions of polarization can be studied by considering the two distinct cases shown in Figure 3.4. The *R!• ne of incidence* is defined as the plane containing the incident, reflected, and transmitted rays fRam65]. In Figure 3.4a, the E-field polarization is parallel with the place of incidence I that is, the E-field has s vertical polarization, or normal component, with respect to the reflecting surface) and in Figure 8.4b, the E-field polarization in perpendicular to the plane of incidence Ithat is, the incident E-field iB printing out of the page toward8 the reader, and ia perpendicular to the page and parallel to the reflecting surface).

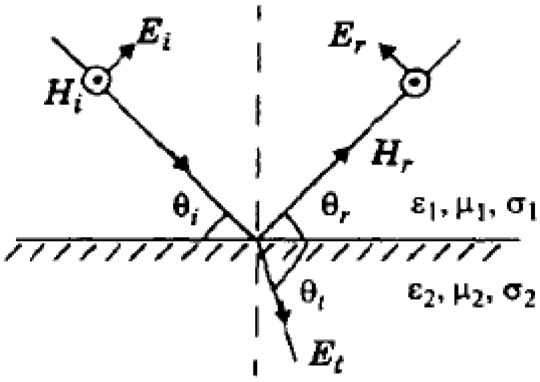
In Figure 8.4, the subscripts i, r. i refer to the incident, rejected, and transmitted fields, respectively. Parameters c, . p,, n, , and c_2, p,, o; represent the permittivity, permeabiiity, and conductance of the two media, respectively. Often, the dielectric constant of a perfect (lossless) dielectric is related to a relative value of permittivity, e, , such that e = etc, , where e_0 la a constant given by 8.85 • 10*" F/rn. If a dielectric material is lousy, it will a1›sorb power and may be described by a complex dielectric constant given by

$$\varepsilon \quad \varepsilon_0 \varepsilon_r - j\varepsilon' \tag{3.17}$$

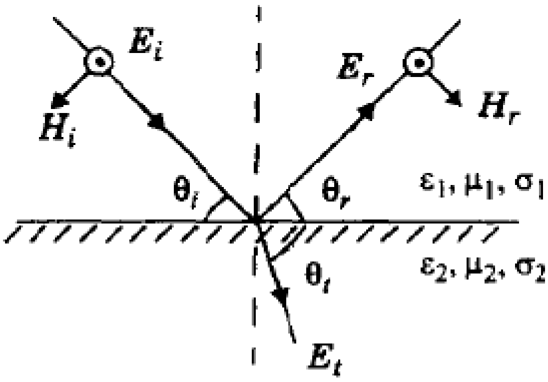
where,

$$\varepsilon' \quad \frac{\sigma}{2\pi f} \tag{3.18}$$

and o is the conductivity of the material measured in Siemens/meter. The terms e, and o are generally insensitive to operating frequency when the material is a good conductor I/<p I e i,) J. For lossy dielectrics, s and i, are generally constant with f<•9• ency, but p may be sensitive to the operating frequency, as shown in Table 8.1. Electrical properties of a wide range of materials were chart acterized over a large frequency range by Von Hippie (Von6z1



la) E-field in the plane of incidence



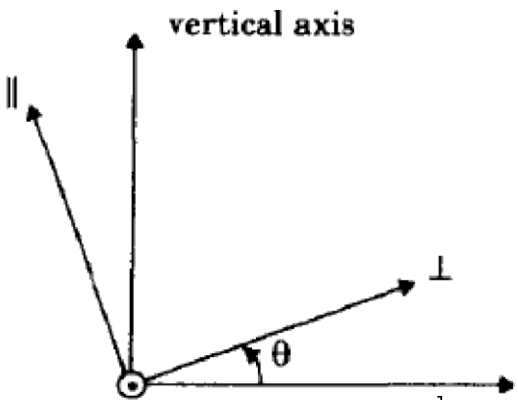
tb) Afield normal to the plane of incidence

Fig ure 3.4
Geometry for cnlculating the reflection rocffi cients between tvrn die lcctri ce

Tablo 3.1 Material Paran›etera at Various Fraqueucies

Material	Relative Permittivily ∘	Conductivity ∘ ‹* m)	Frequency (MHz)
Poor Ground	4	0.001	I00
Typical Ground	1 S	0.005	100
Good Grc›und	25	0.02	I30
Sea Water	81	50	1DU
Fresh lVater		.	
Brick	4.44	0.0U1	4010
Li western	781	0U?0	4000
Glass, Corning 707	4	0OODO0018	1
G.ass, Corning ‘707	4	U000O2T	10L
Glass, Corning 707	4	â0O5	100 0

Recause of superpoc<ition, only tp o orthogonal polarizations need be considered. to snlve general reflection problems. The reflection coefficients for the two cases of parallel and perper.dicu1ar E-field polarization at the hotmdary of two dieluctrics arc given by



$$\Gamma_{!!} \quad \frac{E_r}{E,} \quad \frac{\eta_2 \sin\theta_t - \eta_1 \sin\theta_i}{\bullet| \sin P, + p ,ein 9,}$$ (E-field in pt ane of incidences (S.191

$$\Gamma_{\perp} \quad \frac{E_r}{E_i} \quad \frac{\eta_2 \sin\theta_i - \eta_1 \sin\theta_t}{\eta_2 \sin\theta_i + \eta_1 \sin\theta_t}$$ (E-field not in plane pf incidence1 (fl.201

where p, is the intrinsic impedance of the i th medium I i = 1. 2 1, and is given by i n, , the ratio of electric to magnetic field for a uniform plane wave in the par- ticulw medium. The vel ocity of an electromagnetic wave is given by 1 ' $(\sqrt{\mu\varepsilon})$, and the boundary conditions at thr surface of incidence obey Snell's Law which, referring i,O Figure 8.4, is given by

$$\sqrt{\mu_1 \varepsilon_1} \sin(90 - \theta_i) \qquad \sqrt{\mu_2 \varepsilon_2} \sin(90 - \theta_t) \tag{3.21}$$

The boundary conditions from .Maxwe1l’s equat:ons are used to derive equations (0.19) and (3.20) as well as equations (3.£21, (8.2H. Ei), and (8.28. b).

$$\theta_i \quad \theta_r$$ 13.22)

and

$$E_r \quad \Gamma E_i$$ ‹3.23.a)

$$E_t \quad (1 + \Gamma) E_i$$ t3.23.b\

where P ia either r i, » F , depending on polarization.

For the case when the first medium is free space and p ; p , the re0ection **coefficients for** the two caee8 of vertical rind horizontal polarization **can be** sim- plified to

$$\Gamma_{-} \quad \frac{-\varepsilon_r \sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}}{\varepsilon_r \sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}} \tag{3.24}$$

$$\Gamma_{\perp} \quad \frac{\text{s.n } 8,\text{— p'}\varepsilon_r - \cos^2\theta_i}{\sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}} \tag{3.25}$$

For the case of elliptical p‹iliirized waves, the wave may be broken down (depolarixed) irttn its vertical and horizontal E-field components, rind **superpoSi-** tion may he applied to determine transmitted and reflected waves. In the general case of reflection or transmission, the horizontal and verticnl axes of the spatial coordinates may not coincide with the perpendicular and parallel axes of the propagating waves. An angle 0 measured counter-clockwise from tl:e horizontal axis is defined as shcw'n in Figure 3.5 for a propagating w ave out of the page (towards the reader) [Stu9.3]. ’l’lie vertical and horizontal field components at a dielectric boundary may be related by

Brewster Angle

The *Brtwoter angle is* the angle at which no reflection occurn in the medium of origin. It occurs when the incident angle 8p ia auch that the reflection coeffi- cient P,. is equal to zero (see Figurs 8.6). The Brewster angle is given by the value of 8p which satisfies

$$\sin\ (8g) \quad \sqrt{\frac{\varepsilon_1}{\varepsilon_1 + \varepsilon_2}} \qquad (3.27)$$

For the caae when the firat medium ia free space and the second medium has a relative permittivity e, , equation (3.27) can be expressed aa

$$\sin\ (8g) \quad \frac{\sqrt{\varepsilon_r - 1}}{\sqrt{\varepsilon_r^2 - 1}} \qquad (3.28)$$

Note that the Brewster angle occurs only for vertical (i.e. parallel) {Polarization.

Reffectlon from Perfect **Conductors**

Since electromagnetic energy cannot paaa through a perfect conductor a plane wave incident on a conductor has all of ita energy rejected. A8 the electric field at the surface of the conductor must be equal to zero at all tirne8 in order to obey Maxwells equations, the re0ected wave muat be equal in magnitude to the incident wave. For the caae when E-field polarization is in the plane of incidence, the boundary conditions require that IRamSSl

6, 6, **(3.29)**

and

D; = R, (E-field in plane of incidence) t3.30)

Similarly, for the caae when the E-field is horizontally polarised, the boundary **conditions require** that

$$\theta_i \quad \theta_r \qquad (131)$$

and

A, --L, (E-field not in **plane of incidence)** IN.321

Referring to equations t3.291 **to (3.32),** we see that for a perfect conductor, *r,,* - 1, aa‹1 Fg = -1, r‹•pardless of incident angle. Elliptical polarized waves may be analyzed by using superposition, as sh0Wn in Figurr 3.5 and equation J.26).

Two Ray Reflection Model

Interaction of EM waves with materials having different electrical properties than the material through which the wave is traveling leads to transmitting of energy through the medium and reflection of energy back in the medium of propagation. The amount of energy reflected to the amount of energy incidented is represented by Fresnel reflection coefficient Γ, which depends upon the wave polarization, angle of incidence and frequency of the wave. For example, as the EM waves can not pass through conductors, all the energy is reflected back with angle of incidence equal to the angle of reflection and reflection coefficient Γ = −1. In general, for parallel and perpendicular polarizations, Γ is given by:

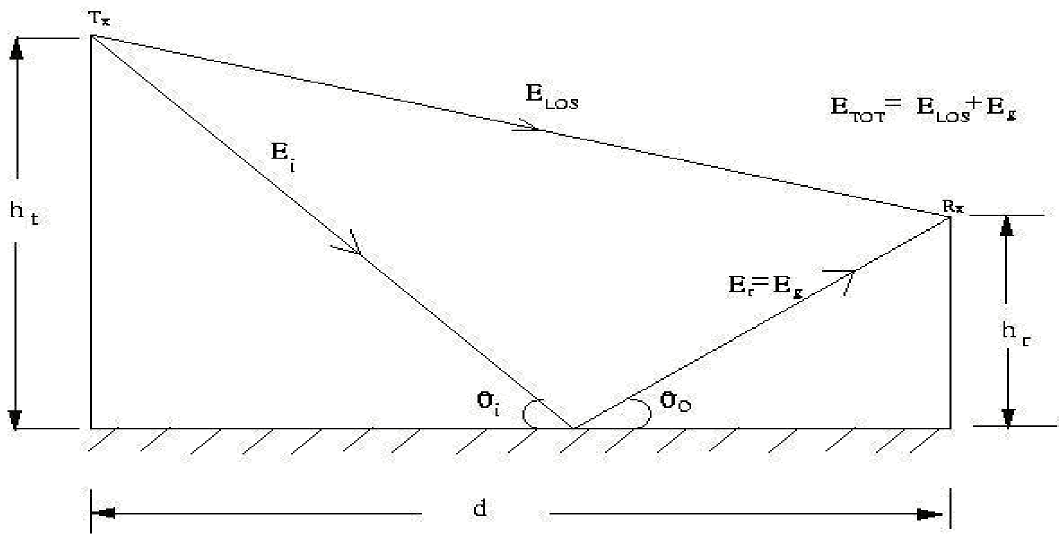


Figure 4.2: Two-ray reflection model.

which agrees very well for large walls made of limestone. The equivalent reflection coefficient is given by,

$$\Gamma_{rough} = \rho_S \Gamma. \quad (4.13)$$

$$\Gamma_{||} = E_r/E_i = \eta_2 \sin \theta_t - \eta_1 \sin \theta_i / \eta_2 \sin \theta_t + \eta_1 \sin \theta_i \qquad (4.14)$$

$$\Gamma_{\perp} = E_r/E_i = \eta_2 \sin \theta_i - \eta_1 \sin \theta_t / \eta_2 \sin \theta_i + \eta_1 \sin \theta_t.$$

Seldom in communication systems we encounter channels with only LOS paths and hence the Friis formula is not a very accurate description of the communication link. A two-ray model, which consists of two overlapping waves at the receiver, one direct path and one reflected wave from the ground gives a more accurate description as shown in Figure 4.2. A simple addition of a single reflected wave shows that power varies inversely with the forth power of the distance between the Tx and the Rx. This is deduced via the following treatment. From Figure 4.2, the total transmitted and received electric fields are

$$E_{TOT}^{T} = E_i + E_{LOS}, \qquad (4.16)$$

$$E_{TOT}^{R} = E_g + E_{LOS}. \qquad (4.17)$$

Let E_0 is the free space electric field (in V/m) at a reference distance d_0. Then

$$E(d, t) = \frac{E_0 d_0}{d} \cos(\omega_c t - \varphi) \qquad (4.18)$$

where

$$\varphi = \omega_c \frac{d}{c} \qquad (4.19)$$

and $d > d_0$. The envelop of the electric field at d meters from the transmitter at any time t is therefore

$$|E(d, t)| = \frac{E_0 d_0}{d}. \qquad (4.20)$$

This means the envelop is constant with respect to time.

Two propagating waves arrive at the receiver, one LOS wave which travels a distance of d' and another ground reflected wave, that travels d''. Mathematically, it can be expressed as:

$$E(d', t) = \frac{E_0 d_0}{d'} \cos(\omega t - \varphi') \tag{4.21}$$

where

$$\varphi' = \omega \frac{d'}{c} \tag{4.22}$$

and

$$E(d'', t) = \frac{E_0 d_0}{d''} \cos(\omega t - \varphi'') \tag{4.23}$$

where

$$\varphi'' = \omega \frac{d''}{c} . \tag{4.24}$$

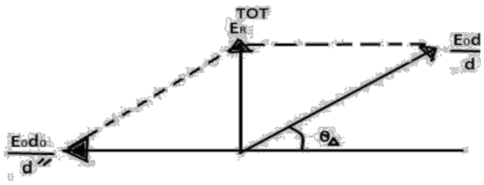


Figure 4.3: Phasor diagram of electric fields.

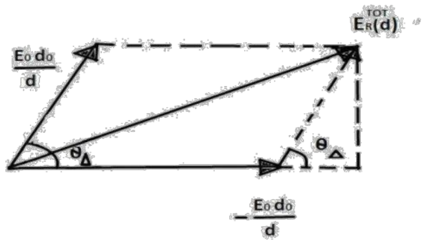


Figure 4.4: Equivalent phasor diagram of Figure 4.3.

According to the law of reflection in a dielectric, $\theta_i = \theta_0$ and $E_g = \Gamma E_i$ which means the total electric field,

$$E_t = E_i + E_g = E_i(1 + \Gamma). \qquad (4.25)$$

For small values of θ_i, reflected wave is equal in magnitude and 180^o out of phase with respect to incident wave. Assuming perfect horizontal electric field polarization, i.e.,

$$\Gamma_\perp = -1 \Longrightarrow E_t = (1 - 1)E_i = 0, \qquad (4.26)$$

the resultant electric field is the vector sum of E_{LOS} and E_g . This implies that,

$$E_R^{TOT} = |E_{LOS} + E_g|. \qquad (4.27)$$

It can be therefore written that

$$E_R^{TOT}(d, t) = \frac{E_0 d_0}{d'} \cos\left(\omega\left(t - \frac{d'}{c}\right)\right) + (-1)\frac{E_0 d_0}{d''} \cos\left(\omega\left(t - \frac{d''}{c}\right)\right) \qquad (4.28)$$

In such cases, the path difference is

$$\Delta = d'' - d' = \sqrt{(h_t + h_r)^2 + d^2} - \sqrt{(h_t - h_r)^2 + d^2}. \qquad (4.30)$$

(4.31)

(4.29)

However, when T-R separation distance is very large compared to ($h_t + h_r$), then

$$\Delta \approx \frac{2h_t h_r}{d}$$

Ex 3: Prove the above two equations, i.e., equation (4.29) and (4.30).

Once the path difference is known, the phase difference is

$$\theta_\Delta = \frac{2\pi\Delta}{\lambda} = \frac{\Delta\omega_c}{c}$$

and the time difference,

$$\tau_d = \frac{\Delta}{c} = \frac{\theta_\Delta}{2\pi f_c}. \qquad (4.32)$$

When d is very large, then Δ becomes very small and therefore E_{LOS} and E_g are virtually identical with only phase difference, i.e.,

$$\left|\frac{E_0 d_0}{d}\right| \approx \left|\frac{E_0 d_0}{d_t}\right| \approx \left|\frac{E_0 d_0}{d_{tt}}\right|. \qquad (4.33)$$

Say, we want to evaluate the received E-field at any $t = \frac{d_t}{c}$. Then,

$$E_{TOT_R}\left(d, t = \frac{d_t}{c}\right) = \frac{E_0 d_0}{d_t}\cos\left(\omega_c\frac{d_t}{c} - \omega_c\frac{d_t}{c}\right) - \frac{E_0 d_0}{d_{tt}}\cos\left(\omega_c\left(\frac{d_t}{c} - \frac{d_{tt}}{c}\right)\right) \qquad (4.34)\text{–}(4.35)$$

$$= \frac{E_0 d_0}{d_t}\cos\left(\frac{\Delta\omega_c}{c}\right) - \frac{E_0 d_0}{d_{tt}}\cos(0^\circ) \qquad (4.36)$$

$$= \frac{E_0 d_0}{d_t}\angle\theta_\Delta - \frac{E_0 d_0}{d_{tt}}$$

$$\approx \frac{E_0 d_0}{d}(\angle\theta_\Delta - 1). \qquad (4.37)$$

Using phasor diagram concept for vector addition as shown in Figures 4.3 and 4.4, we get

$$|E_{TOT_R}(d)| = \sqrt{\left(\frac{E_0 d_0}{d} + \frac{E_0 d_0}{d}\cos(\theta_\Delta)\right)^2 + \left(\frac{E_0 d_0}{d}\sin(\theta_\Delta)\right)^2}$$

$$= \frac{E_0 d_0}{d}\sqrt{(\cos(\theta_\Delta) - 1)^2 + \sin^2(\theta_\Delta)}$$

$$= \frac{E_0 d_0}{d}\sqrt{2 - 2\cos(\theta_\Delta)}$$

$$= 2\frac{E_0 d_0}{d}\sin\left(\frac{\theta_\Delta}{2}\right).$$

For $\frac{\theta_\Delta}{2} < 0.5 rad$, $sin(\frac{\theta_\Delta}{2}) \approx \frac{\theta_\Delta}{2}$. Using equation (4.31) and further equation (4.38) we can then approximate that

$$sin(\frac{\theta_\Delta}{2}) \approx \frac{\pi}{\lambda}\Delta = \frac{2\pi h_t h_r}{\lambda d} < 0.5 rad.$$

(4.39)

(4.40)

This raises the wonderful concept of 'cross-over distance' d_c, defined as (4.41)

$$d > d_c = \frac{20\pi h_t h_r}{5\lambda} = \frac{4\pi h_t h_r}{\lambda}.$$

(4.30),

The corresponding approximate received electric field is

$$E_R^{TOT}(d) \approx 2\frac{E_0 d_0}{d}\frac{2\pi h_t h_r}{\lambda d} = k\frac{h_t h_r}{d^2}.$$

[redacted]

(4.42)

(4.43)

(4.44)

Therefore, using equation (4.43) in (4.1), we get the received power as

$$P_r = \frac{P_t G_t G_r h_t^2 h_r^2}{L d^4}. \quad (4.45)$$

The cross-over distance shows an approximation of the distance after which the received power decays with its fourth order. The basic difference between equation (4.1) and (4.45) is that when $d < d_c$, equation (4.1) is sufficient to calculate the path loss since the two-ray model does not give a good result for a short distance due to the oscillation caused by the constructive and destructive combination of the two rays, but whenever we distance crosses the 'cross-over distance', the power falls off rapidly as well as two-ray model approximation gives better result than Friis equation.

Observations on Equation (4.45): The important observations from this equation are:

1. This equation gives fair results when the T-R separation distance crosses the cross-over distance.

1. In that case, the power decays as the fourth power of distance

$$P_r(d) = \frac{K}{d^4}, \quad (4.46)$$

with K being a constant.

2. Path loss is independent of frequency (wavelength).
3. Received power is also proportional to h_t^2 and h_r^2, meaning, if height of any of the antennas is increased, received power increases.

Diffraction

Diffraction is the phenomena that explain the digression of a wave from a straight line path, under the influence of an obstacle, so as to propagate behind the obstacle. It is an inherent feature of a wave be it longitudinal or transverse. For e.g the sound can be heard in a room, where the source of the sound is another room without having any line of sight. The similar phenomenon occurs for light also but the diffracted light intensity is not noticeable. This is because the obstacle or slit need to be of the order of the wavelength of the wave to have a significant effect. Thus radiation from a point source radiating in all directions can be received at any

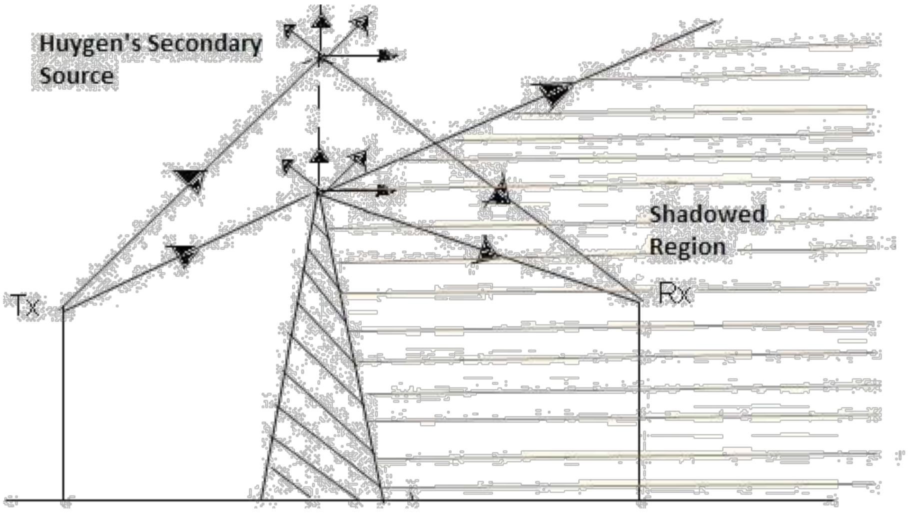


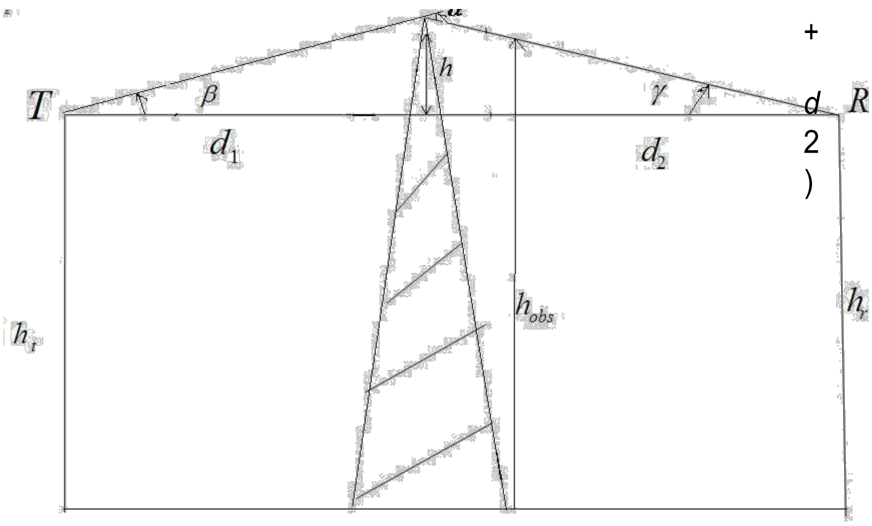
Figure 4.5: Huygen's secondary wavelets.

point, even behind an obstacle (unless it is not completely enveloped by it), as shown in Figure 4.5. Though the intensity received gets smaller as receiver is moved into the shadowed region. Diffraction is explained by Huygens-Fresnel principle which states that all points on a wavefront can be considered as the point source for secondary wavelets which form the secondary wavefront in the direction of the prorogation. Normally, in absence of an obstacle, the sum of all wave sources is zero at a point not in the direct path of the wave and thus the wave travels in the straight line. But in the case of an obstacle, the effect of wave source behind the obstacle cannot be felt and the sources around the obstacle contribute to the secondary wavelets in the shadowed region, leading to bending of wave. In mobile communication, this has a great advantage since, by diffraction (and scattering, reflection), the receiver is able to receive the signal even when not in line of sight of the transmitter. This we show in the subsection given below.

Knife-Edge Diffraction Geometry

As shown in Figure 4.6, consider that there's an impenetrable obstruction of hight h at a distance of $d1$ from the transmitter and $d2$ from the receiver. The path difference between direct path and the diffracted path is

$$\delta = \sqrt{\quad 2 + h2 +} \sqrt{d^2 + h^2} - (d1$$



\+ d 2) (4.47)

Figure 4.6: Diffraction through a sharp edge.

which can be further simplified as

$$\delta = d_1(1 + h^2/2d_1^2) + d_2(1 + h^2/2d_2^2) - (d_1 + d_2)$$

$$= h^2/(2d_1) + h^2/(2d_2) = h^2(d_1 + d_2)/(2d_1d_2). \quad (4.48)$$

Thus the phase difference equals

$$\varphi = 2\pi\delta/\lambda = 2\pi h^2(d_1 + d_2)/\lambda 2(d_1d_2). \quad (4.49)$$

With the following considerations that

$$\alpha = \beta + \gamma \quad (4.50)$$

and

$$\alpha \approx \tan\alpha \quad (4.51)$$

we can write,

$$\alpha\tan\alpha = \tan\beta + \tan\gamma = h/d_1 + h/d_2 = h(d_1 + d_2)/d_1d_2. \quad (4.52)$$

In order to normalize this, we usually use a Fresnel-Kirchoff diffraction parameter v, expressed as

$$v = h\sqrt{2(d_1 + d_2)/(\lambda d_1d_2)} = \alpha\sqrt{(2d_1d_2)/(\lambda(d_1 + d_2))} \quad (4.53)$$

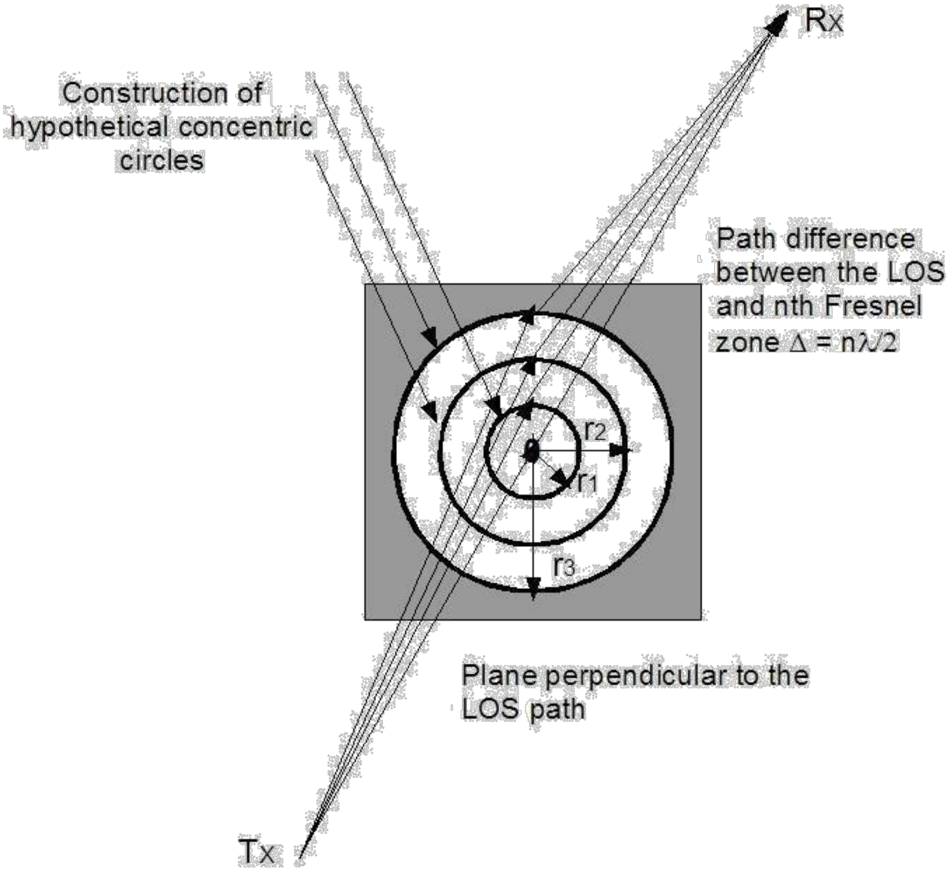


Figure 4.7: Fresnel zones.

and therefore the phase difference becomes

$$\varphi = \pi v^2/2. \tag{4.54}$$

From this, we can observe that: (i) phase difference is a function of the height of the obstruction, and also, (ii) phase difference is a function of the position of the obstruction from transmitter and receiver.

Fresnel Zones: the Concept of Diffraction Loss

As mentioned before, the more is the object in the shadowed region greater is the diffraction loss of the signal. The effect of diffraction loss is explained by Fresnel zones as a function of the path difference. The successive Fresnel zones are limited by the circular periphery through which the path difference of the secondary waves is $n\lambda/2$ greater than total length of the LOS path, as shown in

Figure 4.7. Thus successive Fresnel zones have phase difference of π which means they alternatively

provide constructive and destructive interference to the received the signal. The radius of the each Fresnel zone is maximum at middle of transmitter and receiver (i.e. when $d1 = d2$) and decreases as moved to either side. It is seen that the loci of a Fresnel zone varied over d1 and d2 forms an ellipsoid with the transmitter and receiver at its focii. Now, if there's no obstruction, then all Fresnel zones result in only the direct LOS prorogation and no diffraction effects are observed. But if an obstruction is present, depending on its geometry, it obstructs contribution from some of the secondary wavelets, resulting in diffraction and also the loss of energy, which is the vector sum of energy from unobstructed sources. please note that height of the obstruction can be positive zero and negative also. The diffraction losses are minimum as long as obstruction doesn't block volume of the 1st Fresnel zone. As a rule of thumb, diffraction effects are negligible beyond 55% of 1st Fresnel zone.

Ex 4: Calculate the first Fresnel zone obstruction height maximum for f = 800 MHz.

Solution:

$$\lambda = \frac{c}{f} = \frac{3 \times 10^8}{8 \times 10^2 \times 10^6} = \frac{3}{8} m$$

$$H = \sqrt{\frac{\lambda\, d1\, d2}{(d1 + d2)}}$$

$$H1 = \sqrt{\frac{\frac{3}{8}\, 250 \times 250}{500}} = 6.89m$$

Thus H1 = 10 + 6.89 = 16.89m

(b)

$$H2 = \sqrt{\frac{\frac{3}{8} \times 100 \times 400}{500}} = 10\,(0.3) = 5.48m$$

Thus

$$H2 = 10 + 5.6 = 15.48m$$

. To have good power strength, obstacle should be within the 60% of the first fresnel zone.

Ex 5: Given f=900 MHz, $d1 = d2 = 1$ km, h = 25m, where symbols have usual meaning. Compute the diffraction loss. Also find out in which Fresnel zone the tip of the obstruction lies.

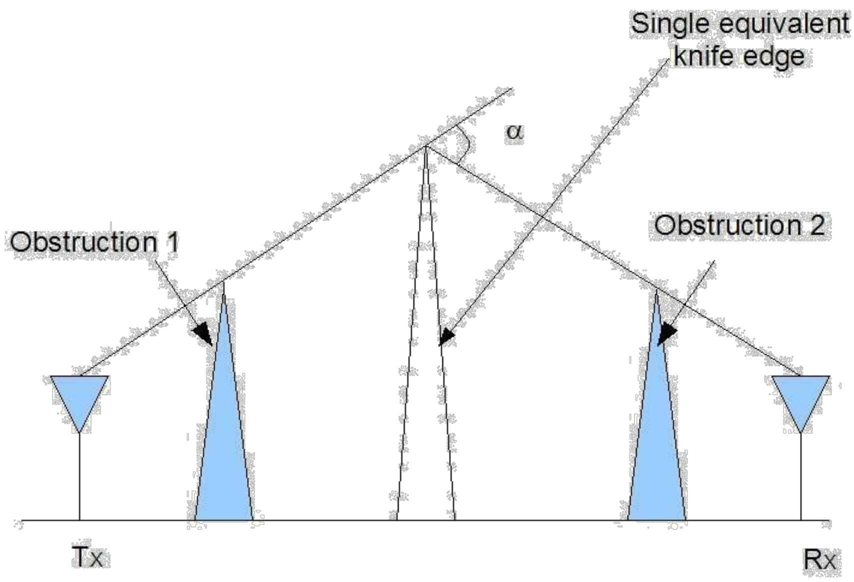


Figure 4.8: Knife-edge Diffraction Model

Given,

$$G_d(dB) = 20\log(0.5 - 0.62v) \qquad -1 < v <= 0$$

$$G_d(dB) = 20\log(0.225/v) \qquad v > 2.24$$

Solution:

$$v = h\sqrt{\frac{2(d_1 + d_2)}{\lambda d_1 d_2}} = 25\sqrt{\frac{2 \times 2000}{\frac{1}{3}\ 10}} = 2.74$$

$$G_d(dB) = 20\log\left(\frac{0.225}{v}\right) = -21.7dB$$

Since loss = $-G_d$ (dB) = 21.7 dB

$$n = \frac{(2.74)^2}{2} = 3.5$$

Thus n=4.

4.5.3 Knife-edge diffraction model

Knife-edge diffraction model is one of the simplest diffraction model to estimate the diffraction loss. It considers the object like hill or mountain as a knife edge sharp object. The electric field strength, E_d of a knife-edge diffracted wave is given by

$$E_d/E_o = F(v) = (1+j)/2 \int_{v}^{\infty} (\exp((-j\pi t^2)/2)dt \quad (4.55)$$

The diffraction gain due to presence of knife edge can be given as

$$G_d(db) = 20\log|F(v)| \quad (4.56)$$

$$G_d(db) = 0 \quad v <= -1 \quad (4.57)$$

$$G_d(db) = 20\log(0.5 - 0.62) \quad -1 <= v <= 0 \quad (4.58)$$

$$G_d(db) = 20\log(0.5\exp(-0.95v)) \quad 0 <= v <= 1 \quad (4.59)$$

$$G_d(db) = 20\log(0.4 - sqrt(0.1184 - (0.38 - 0.1v^2))) \quad 1 <= v <= 2.4 \quad (4.60)$$

$$G_d(db) = 20\log(0.225/v) \quad v > 2.4 \quad (4.61)$$

When there are more than one obstruction, then the equivalent model can be found by one knife-edge diffraction model as shown in Figure 4.8.

Scattering

The actual received power at the receiver is somewhat stronger than claimed by the models of reflection and diffraction. The cause is that the trees, buildings and lamp-posts scatter energy in all directions. This provides extra energy at the receiver. Roughness is tested by a Rayleigh criterion, which defines a critical height h_c of surface protuberances for a given angle of incidence θ_i, given by,

$$h_c = \frac{\lambda}{8\sin\theta_i}. \quad (4.10)$$

A surface is smooth if its minimum to maximum protuberance h is less than h_c, and rough if protuberance is greater than h_c. In case of rough surfaces, the surface reflection coefficient needs to be multiplied by a scattering loss factor ρ_S, given by

$$\left(\frac{\pi\sigma_h \sin\theta_i}{\quad}\right)^2) \quad (4.11)$$

$$\rho_S = exp(-8(\frac{\quad}{\lambda}$$

where σ_h is the standard deviation of the Gaussian random variable h. The following result is a better approximation to the observed value

$$\rho_S = exp\left(-8\left(\frac{\pi\sigma_h sin\theta_i}{\lambda}\right)^2\right) I_0\left[-8\left(\frac{\pi\sigma_h sin\theta_i}{\lambda}\right)^2\right] \quad (4.12)$$

Outdoor Propagation Models

There are many empirical outdoor propagation models such as Longley-Rice model, Durkin's model, Okumura model, Hata model etc. Longley-Rice model is the most commonly used model within a frequency band of 40 MHz to 100 GHz over different terrains. Certain modifications over the rudimentary model like an extra urban factor (UF) due to urban clutter near the reciever is also included in this model. Below, we discuss some of the outdoor models, followed by a few indoor models too.

3.10.1 Longley-Rice Model

The Longley-Rice model [Ric67], [Lon68] is applicable to point-to-point communication systems in the frequency range from 40 MHz to 100 GHz, over different kinds of terrain. The median transmission loss is predicted using the path geometry of the terrain profile and the refractivity of the troposphere. Geometric optics techniques (primarily the 2-ray ground reflection model) are used to predict signal strengths within the radio horizon. Diffraction losses over isolated obstacles are estimated using the Fresnel-Kirchoff knife-edge models. Forward scatter theory is used to make troposcatter predictions over long distances, and far field diffraction losses in double horizon paths are predicted using a modified Van der Pol-Bremmer method. The Longley-Rice propagation prediction model is also referred to as the *ITS irregular terrain model*.

The Longley-Rice model is also available as a computer program [Lon78] to calculate large-scale median transmission loss relative to free space loss over irregular terrain for frequencies between 20 MHz and 10 GHz. For a given transmission path, the program takes as its input the transmission frequency, path length, polarization, antenna heights, surface refractivity, effective radius of earth, ground conductivity, ground dielectric constant, and climate. The program also operates on path-specific parameters such as horizon distance of the antennas, horizon elevation angle, angular trans-horizon distance, terrain irregularity and other specific inputs.

The Longley-Rice method operates in two modes. When a detailed terrain path profile is available, the path-specific parameters can be easily determined and the prediction is called a *point-to-point mode* prediction. On the other hand, if the terrain path profile is not available, the Longley-Rice method provides techniques to estimate the path-specific parameters, and such a prediction is called an *area mode* prediction.

There have been many modifications and corrections to the Longley-Rice model since its original publication. One important modification [Lon78] deals with radio propagation in urban areas, and this is particularly relevant to mobile radio. This modification introduces an excess term as an allowance for the additional attenuation due to urban clutter near the receiving antenna. This extra term, called the *urban factor* (UF), has been derived by comparing the predictions by the original Longley-Rice model with those obtained by Okumura [Oku68].

One shortcoming of the Longley-Rice model is that it does not provide a way of determining corrections due to environmental factors in the immediate vicinity of the mobile receiver, or consider correction factors to account for the effects of buildings and foliage. Further, multipath is not considered.

Okumura Model

The Okumura model is used for Urban Areas is a Radio propagation model that is used for signal prediction.The frequency coverage of this model is in the range of 200 MHz to 1900 MHz and distances of 1 Km to 100 Km.It can be applicable for base station effective antenna heights (*ht*) ranging from 30 m to 1000 m.

Okumura used extensive measurements of base station-to-mobile signal attenua-tion throughout Tokyo to develop a set of curves giving median attenuation relative to free space (*Amu*) of signal propogation in irregular terrain. The empirical path-loss formula of Okumura at distance *d* parameterized by the carrier frequency *fc* is given by

$$PL(d)dB = L(fc, d) + Amu(fc, d) - tt(ht) - tt(hr) - ttAREA \qquad (4.67)$$

where *L*(*fc, d*) is free space path loss at distance d and carrier frequency *fc*, *Amu*(*fc, d*) is the median attenuation in addition to free-space path loss across all environments, *tt*(*ht*) is the base station antenna height gain factor, *tt*(*hr*) is the mobile antenna height gain factor, *ttAREA* is the gain due to type of environment. The values of *Amu*(*fc, d*) and *ttAREA* are obtained from Okumura's empirical plots. Okumura derived empirical formulas for *tt*(*ht*) and *tt*(*hr*) as follows:

$$tt(ht) = 20\log_{10}(ht/200), \quad 30m < ht < 1000m \qquad (4.68)$$

$$tt(hr) = 10\log_{10}(hr/3), \quad hr \leq 3m \qquad (4.69)$$

$$tt(hr) = 20\log_{10}(hr/3), \quad 3m < hr < 10m \qquad (4.70)$$

Correlation factors related to terrain are also developed in order to improve the models accuracy. Okumura's model has a 10-14 dB empirical standard deviation between the path loss predicted by the model and the path loss associated with one of the measurements used to develop the model.

Hata Model

The Hata model is an empirical formulation of the graphical path-loss data provided by the Okumura and is valid over roughly the same range of frequencies, 150-1500 MHz. This empirical formula simplifies the calculation of path loss because it is closed form formula and it is not based on empirical curves for the different param-eters. The standard formula for empirical path loss in urban areas under the Hata model is

$$P_{L,urban}(d)dB = 69.55 + 26.16\log_{10}(fc) - 13.82\log_{10}(ht) - a(hr) + (44.9 - 6.55\log_{10}(ht))\log_{10}(d)$$

The parameters in this model are same as in the Okumura model,and $a(hr)$ is a correction factor for the mobile antenna height based on the size of coverage area.For small to medium sized cities this factor is given by

$$a(hr) = (1.11 \log_{10}(fc) - 0.7)hr - (1.56 \log_{10}(fc) - 0.8)dB$$

and for larger cities at a frequencies $fc > 300$ MHz by

$$a(hr) = 3.2(\log_{10}(11.75hr))^2 - 4.97dB$$

else it is

$$a(hr) = 8.29(\log_{10}(1.54hr))^2 - 1.1dB$$

Corrections to the urban model are made for the suburban, and is given by

$$PL, suburban(d)dB = PL, urban(d)dB - 2(\log10(fc/28))^2 - 5.4 \quad (4.72)$$

Unlike the Okumura model, the Hata model does not provide for any specific path-correlation factors. The Hata model well approximates the Okumura model for distances d > 1 Km. Hence it is a good model for first generation cellular systems, but it does not model propogation well in current cellular systems with smaller cell sizes and higher frequencies. Indoor environments are also not captured by the Hata model.

PCS Extension to lata Model

The European Co-operative for SCientific and' Technical research IEURO-COST) formed the COST-231 working committee to develop an extended ›-ersion of the Hata model. COST-231 proposed the following formula to extend Hata's model to 2 GHG. The proposed model for path loss is {EUR91)

$$L_{50}(urban) = 46.3 + 33.9\log f_c - 13.82\log h_{te} - a(h_{re}) + (44.9 - 6.55\log h_{te})\log d + C_M \quad (3.87)$$

where is defined in equations f3.83), i3.84.a), and (3.84.b i and

$$C = \begin{cases} 0\ \text{dB} & \text{fur medium sized city and suburban areas} \\ \text{dB} & \text{for metropolitan centers} \end{cases} \quad (3.88)$$

The COST-281 extension of the Hata model is restricted to the following range of parameters:

: 1500 MHz to 2000 MHz

h,, : 80 m to 2D0 m

h,, : \ rn to 10 m

d : \ km to 20 km

Walflsch and Bertoni Model

A model developed by Walfisch and Bertoni [Wal88l tonaiders the impact of rooftops and building height by using diffraction to predict average signal strength at street level. The model considers the path lose, S:, to be a product of three factors.

$$S = P_0 Q^2 P_1 \quad (3.89)$$

wherer P_0 repre6ents free space path losa between isotropic antennas given by

$$P_0 = \left(\frac{\lambda}{4\pi R}\right)^2 \quad (3.90)$$

The factor Q gives the reduction in the rooftop signal due to the rnw of buildings whith immediately shadow the receiver at street level. The P, term is t›a»ed upon diffraction and de terminus tnc signal los.s front the.: rooftop to the street.

Ip dH, thi• p:ith ltd-s is given by

$$S(\text{dB}) \quad L_0 + L_{rts} + L_{ms} \qquad (3.91)$$

where L_a represents free space losa, L„, represents the "rooftop-to-street diffraction and scatter loss", and *Lq* denotes rDultiscreer diffraction loss due to the *Fowl* of hiiil drugs aS2]. Figure .4.ES illustrams the geometry used in the Wiiifisuh Bertoni model *(Wa I R R I,* fMar93l, This model is heinp cons.dered for use by ITU-R in the 1bI'l'-2000 s idards activities.

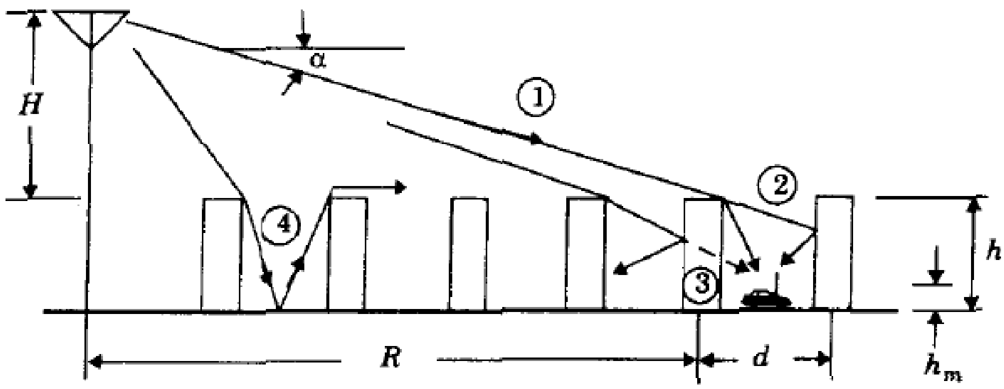


Figure 8.25
Propngotion geometry for mode. proposed hy Welfisch and Bertnni [l"mm lh'a1S8l H' iE EE].

Wideband PCS MiCrOceII Modal

Work by Feuerstein, et.aI. in US 1 used a 20 MHZ pulsed transmitter at 1900 MHz to measure put h loss, outage, and delay spread in typical microceEiilar systems in San l"rancisco and Oakland. Using hose station antenna heights of 3.7 m, 8. 3 m, rind 13. d m, and a mobile receiver with on antenna heigb t of 1.7 m above ground, statistics for p:ith loss, multipath, and coverage area w'ere developed from extensive meauu resents in line-of-sight 'LCifi) and *obai*ru *cm d* l UBS) en vi ronments (Feii94l This work revealed that a 2-ray gruund Fef)ectit›n model (shown in I•"igli rt• S.7 is a good estimate for path loss in LOS micCucells, ind a s inn ple log-distance path loss model hold.s 'till for OBS microcell environments.

For a flat earth gvcund reflection model, the distance *d,* st wh:ch the fire- t -"res nel zone just becomes ob structcd by the ground Ifir9t Frcsncl zorie cleRrince) is given by

$$d_f = \frac{1}{\lambda}\sqrt{(\Sigma^2 - \Delta^2)^2 - 2(\Sigma^2 + \Delta^2)\left(\frac{\lambda}{2}\right)^2 + \left(\frac{\lambda}{2}\right)^4} \quad (3.92.a)$$

$$= \frac{1}{\lambda}\sqrt{16h_t^2h_r^2 - \lambda^2\left(h_t^2 + h_r^2\right) + \frac{\lambda^4}{16}}$$

For LOS cases, a double regression path loss model that uses a regression breakpoint at the first Fresnel zone clearance was shown to fit well to measurements. The model assumes omnidirectional vertical antennas and predicts average path loss as

$$PL(d) = \begin{cases} 10n_1 \log(d) + p_1 & \text{for } 1 < d < d_f \\ 10n_2 \log(d/d_f) + 10n_1 \log d_f + p_1 & \text{for } d > d_f \end{cases} \quad (3.92.b)$$

where p_1 is equal to $\overline{PL}(d_0)$ (the path loss in decibels at the reference distance of $d_0 = 1$ m), d is in meters and n_1, n_2 are path loss exponents which are a function of transmitter height, as given in Figure 3.26. It can easily be shown that at 1900 MHz, $p_1 = 38.0$ dB.

For the OBS case, the path loss was found to fit the standard log-distance path loss law of equation (3.69.a)

$$\overline{PL}(d)\,[dB] = 10n\log(d) + p_1 \quad (3.92.c)$$

where n is the OBS path loss exponent given in Figure 3.26 as a function of transmitter height. The standard deviation (in dB) of the log-normal shadowing component about the distance-dependent mean was found from measurements using the techniques described in Chapter 3, section 3.10.2. The log-normal shadowing component is also listed as a function of height for both the LOS and OBS microcell environments. Figure 3.26 indicates that the log-normal shadowing component is between 7 and 9 dB regardless of antenna height. It can be seen that LOS environments provide slightly less path loss than the theoretical 2-ray ground reflected model, which would predict $n_1 = 2$ and $n_2 = 4$.

Transmitter Antenna Height	1900 MHz LOS n_1	n_2	σ(dB)	1900 MHz OBS n	σ(dB)
Low (3.7m)	2.18	3.29	8.76	2.58	9.31
Medium (8.5m)	2.17	3.36	7.88	2.56	7.67
High (13.3m)	2.07	4.16	8.77	2.69	7.94

Figure 3.26

Indoor Propagation Models

The indoor radio channel differs from the traditional mobile radio channel in ways - the distances covered are much smaller ,and the variability of the environment is much greater for smaller range of Tx-Rx separation distances.Features such as lay-out of the building,the construction materials,and the building type strongly in-fluence the propagation within the building.Indoor radio propagation is dominated by the same mechanisms as outdoor: reflection, diffraction and scattering with vari-able conditions. In general,indoor channels may be classified as either line-of-sight or obstructed.

Partition Losses Inside a Floor (Intra-floor)

The internal and external structure of a building formed by partitions and obstacles vary widely.Partitions that are formed as a part of building structure are called hard partitions , and partitions that may be moved and which do not span to the ceiling are called soft partitions. Partitions vary widely in their physical and electrical characteristics,making it difficult to apply general models to specific indoor installations.

Partition Losses Between Floors (Inter-floor)

The losses between floors of a building are determined by the external dimensions and materials of the building,as well as the type of construction used to create the floors and the external surroundings. Even the number of windows in a building and the presence of tinting can impact the loss between floors.

Log-distance Path Loss Model

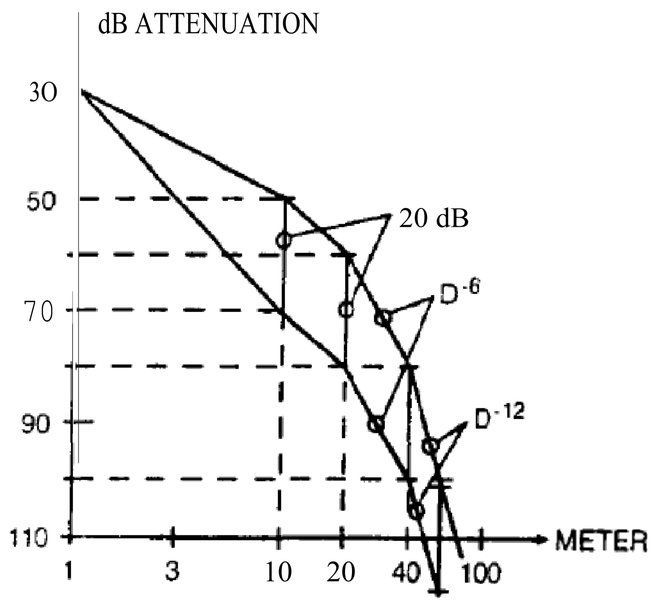
It has been observed that indoor path loss obeys the distance power law given by

$$PL(dB) = PL(d0) + 10n \log_{10}(d/d0) + X\sigma \quad (4.73)$$

where n depends on the building and surrounding type, and $X\sigma$ represents a normal random variable in dB having standard deviation of σ dB.

Ericsson Multiple Breakpoint Model

The Ericsson radio system model was obtained by measurements in a multiple *Roor* c›ffi*i:e* building {Ake88]. The model has four breakpoints and considers both an upper and lower bound on the path loss. The model also assumes that there is SO dB atmnua(ion at. *dq* ‘- *!* m, which can be Bhown to be accurate for / • 900 MHz and unity gain antennas. Rather than assuming a log-normal shadowing component. the Ericsson model provides a deterministic limit on the range of path loas at a particular distance Bernhardt lBer89) used a uniform distribution to generate path loas values within the maximum and minimum range as a function of distance for in-building simulation. Figure 3.Z7 showa a plot of in-building path lose based on the Ericsson model as n function of dis-



Attenuation **Factor Model**

An in-building propagation model that included the effect of building t j›e as well aa the variations caused by obstacles wa8 described by Seidel (Sei92b). This model provides flexibility and waG shown to reduce the standard deviation between measured and predicted path lo88 to around 4 dB, as compared to 13 dB when only a log-distance model was used in two different buildings. The attenuation factor model is given by floprs, it is difficuJt to determine exact mcdels for penetration as only a limited number of experiments hive been published, and they are enznetirnex difficult to compare. However, some generalizations can be made fmm the literature. to measurements reported to date, signal strength received inaide a building increases with height. At the lower floora of a building, the urban clutter induces **greater** attenuation and reduces the level of penetration. At higher Ooors, a LOS path may exist, thus causing a stronger incident signal at, the. exterior wall of the hii ildirin.

*PLtd)l*d (doll)+10nsr*os(„;₀)+* *AFi*d*I t8.94J

where $n_{S\ F}$ represents the exponent value for the “same floor” measurement. Thus, if a good estimate for n exista ie.g., selected from Table 3.4 or Table 8.61 on the same floor, then the path loss on a different floor can be predicted by adding an appropriate value of FAF (e.g„ selected from Table 3.51 Alternatively, in equation (3.941, FAF may be replaced by an exponent which at ready considers the effects of multiple floor separation.

$$\overline{PL}(d)[\mathrm{dB}] \quad \overline{PL}(d_0) + 10n_{MF}\log\left(\frac{d}{d_0}\right) \qquad (3.95)$$

where app denotes a path losa exponent based on measurements through multiple floors.

Table 3.7 illustrates typical values of n for a wide range of locations in many buildings. This table also illustrates how the standard deviation decreases ae the average region becomes smaller and more site specific. Scatter plots illustrating actual measured path loss in two multi-floored office buildings are shown in Firure 8.28 and Fivvtrr fl.29.

RF penetration has been foumd to be a function of frequency as well as height within the building. The antenna pattern in the eleva£inn plane aleo plays an important role in how much signal penetrates a building from the outside. Most measurements hsve considered outdoor transmitters with antenna heights far less than the maximum height of the building under test. Measurements in Liverpool tTars7) showed that penetration loss decreases with increasing frequency. Specifically, penetration attenuation values of 16.4 dB, 11.6 dB, and 7.6 dB were mesaured on the ground moor of a building at frequencies of 441 MHz, 896.5 MHz, and 1400 MHz, respectively. MeaBurements by Turkmani ITur92) showed penetration loss of 14.2 dB, 13.4 , and 12.8 dB for 900 MHx, 1800 MHz, and 2300 MHz, respectively. Measurements made in front of windows indicated 6 dB leaa penetration loss on average than did meaBureznents m8de in parts of the buildings without windows.

3.13 Ray Tracing and Site Specific Modeling

In recent years, the computational and visualization capabilities of computers have accelerated rapidly. New methods for predicting radio oigna) •o erage involve the uee of SIte SPecific (SUP) propagation models and *graphical inform*

zaotionp.stem(GIS)databaaes}Rus93lSISPmodelsBuppomzaytraWng as means of determiniatically modeling any indoor or outdoor propagation environment. ’l’hrnugh the use of building databases, which may be drawn or digitized using standard graphical software packages, wireleoa ayatem designers are able tn include accurate representations of building and terrain features.

For outdoor propagation prediction, ray tracing techniques are uaed in conjunction with aerial photographs 8n that three-dimensional (3-DI repreoentatior s of buildings may be integrated with software that tarriea out rejection, diffraction, and scattering rriodels. Photogrammetrip techniques are used to convert aerial or satellite photographs of cities into usable S-D databases for the models (Sch92], [Itos93J, (Wag84l. In indoor environnsente, architectural drawings provide a aide specific representation for propagation models (Val93), (Sei94), (Kre94).

in nuiiamg eat baaea become prevalent, wireless systems will be developed using computer aided design teols that provide deterministic, rather than statistical, prediction modela for large-scale path loas in a wide range of operatrig environtrients

Walker }Waiszj measured radio signals into fourteen different buildings in Chicago from aeven external cellular transmitters. Results showed that building penetration loas decreased at a rate of 1.9 dB per door from the groimd level up to the Cfteenth door and then began increasing above the fifteenth door. The increase in penetration loss at higher doora we attributed to shadowing effects of adjacent buildings. Similarly, Thrl nani f'f\ir87l reported penetration loas decreased at a rate of 2 dB per troop from the ground level up to the ninth Ooar and then iricreaaed above the ninth door. Similar results were also reported by Durante (Dur78).

Measurements have ahowri that the percentage of windows. when compared with the building face surface area, impacts the level of RF penetration loss, as does tlte presence of tinted metal in the windows. Metallic tints can provide from 3 dB to 30 dB RF attenuation in a single pane of glass. The angle of incidence of the tranBmitted wave upon the face of the building also has a strong impact on the penetration loss, ae was shown by Horiicishi (Hor86).

Unit-3 Mobile Radio Propagation: Small-Scale Fading and Multipath

Multipath Propagation

In wireless telecommunications, multipath is the propagation phenomenon that results in radio signals reaching the receiving antenna by two or more paths. Causes of multipath include atmospheric ducting, ionospheric reflection and refraction, and reflection from water bodies and terrestrial objects such as mountains and buildings. The effects of multipath include constructive and destructive interference, and phase shifting of the signal. In digital radio communications (such as GSM) multipath can cause errors and affect the quality of communications. We discuss all the related issues in this chapter.

Multipath & Small-Scale Fading

Multipath signals are received in a terrestrial environment, i.e., where different forms of propagation are present and the signals arrive at the receiver from transmitter via a variety of paths. Therefore there would be multipath interference, causing multi-path fading. Adding the effect of movement of either Tx or Rx or the surrounding clutter to it, the received overall signal amplitude or phase changes over a small amount of time. Mainly this causes the fading.

Fading

The term **fading**, or, small-scale fading, means rapid fluctuations of the amplitudes, phases, or multipath delays of a radio signal over a short period or short travel distance. This might be so severe that large scale radio propagation loss effects might be ignored.

Multipath Fading Effects

In principle, the following are the main multipath effects:

1. Rapid changes in signal strength over a small travel distance or time interval.
2. Random frequency modulation due to varying Doppler shifts on different mul-tipath signals.
3. Time dispersion or echoes caused by multipath propagation delays.

Factors Influencing Fading

The following physical factors influence small-scale fading in the radio propagation channel:

(1) Multipath propagation – Multipath is the propagation phenomenon that re-sults in radio signals reaching the receiving antenna by two or more paths. The effects of multipath include constructive and destructive interference, and phase shifting of the signal.

(2) Speed of the mobile – The relative motion between the base station and the mobile results in random frequency modulation due to different doppler shifts on each of the multipath components.

(3) Speed of surrounding objects – If objects in the radio channel are in mo-tion, they induce a time varying Doppler shift on multipath components. If the surrounding objects move at a greater rate than the mobile, then this effect dominates fading.

(4) Transmission Bandwidth of the signal – If the transmitted radio signal bandwidth is greater than the "bandwidth" of the multipath channel (quanti-fied by coherence bandwidth), the received signal will be distorted.

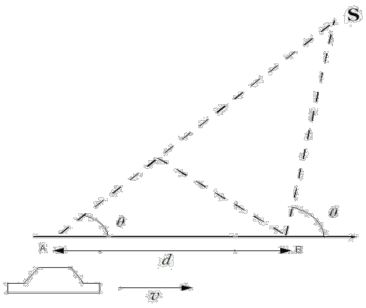


Figure 5.1: Illustration of Doppler effect.

Doppler Shift

The Doppler effect (or Doppler shift) is the change in frequency of a wave for an observer moving relative to the source of the wave. In classical physics (waves in a medium), the relationship between the observed frequency f and the emitted frequency f_0 is given by:

$$f = \frac{v \pm v_r}{v \pm v_s} f_0 \qquad (5.10)$$

where v is the velocity of waves in the medium, v_s is the velocity of the source relative to the medium and v_r is the velocity of the receiver relative to the medium.

In mobile communication, the above equation can be slightly changed according to our convenience since the source (BS) is fixed and located at a remote elevated level from ground. The expected Doppler shift of the EM wave then comes out to be $\pm \frac{v}{c} f_0$ or, $\pm \frac{v_r}{\lambda}$. As the BS is located at an elevated place, a cos φ factor would also be multiplied with this. The exact scenario, as given in Figure 5.1, is illustrated below.

Consider a mobile moving at a constant velocity v, along a path segment length d between points A and B, while it receives signals from a remote BS source S. The difference in path lengths traveled by the wave from source S to the mobile at points A and B is $\Delta l = d \cos\theta = v\Delta t \cos\theta$, where Δt is the time required for the mobile to travel from A to B, and θ is assumed to be the same at points A and B since the source is assumed to be very far away. The phase change in the received signal due to the difference in path lengths is therefore

$$\Delta\phi = \frac{2\pi\Delta l}{\lambda} = \frac{2\pi v\Delta t}{\lambda}\cos\theta \qquad (5.11)$$

and hence the apparent change in frequency, or Doppler shift (f_d) is

$$f_d = \frac{1}{2\pi}\cdot\frac{\Delta\phi}{\Delta t} = \frac{v}{\lambda}\cdot\cos\theta. \qquad (5.12)$$

Example 1

An aircraft is heading towards a control tower with 500 kmph, at an elevation of 20°. Communication between aircraft and control tower occurs at 900 MHz. Find out the expected Doppler shift.

Solution As given here,

$$v = 500 \text{ kmph}$$

the horizontal component of the velocity is

$$v^t = v \cos\theta = 500 \times \cos 20^\circ = 130 \text{ m/s}$$

Hence, it can be written that

$$\lambda = \frac{900 \times 10^6}{3 \times 10^8} = \frac{1}{3} \text{ m}$$

$$f_d = \frac{130}{1/3} = 390 \text{ Hz}$$

If the plane banks suddenly and heads for other direction, the Doppler shift change will be 390 Hz to −390 Hz.

Impulse Response Model of a Multipath Channel

Mobile radio channel may be modeled as a linear filter with time varying impulse response in continuous time. To show this, consider time variation due to receiver motion and time varying impulse response $h(d, t)$ and $x(t)$, the transmitted signal. The received signal $y(d, t)$ at any position d would be

$$y(d, t) = x(t) * h(d, t) = \int_{-\infty}^{\infty} x(\tau)\, h(d, t - \tau)\, d\tau \quad (5.13)$$

For a causal system: $h(d, t) = 0$, for $t < 0$ and for a stable system $\int_{-\infty}^{\infty} |h(d, t)|\, dt < \infty$

Applying causality condition in the above equation, $h(d, t-\tau) = 0$ for $t-\tau < 0 \Rightarrow \tau > t$, i.e., the integral limits are changed to

$$y(d, t) = \int_{-\infty}^{t} x(\tau)\, h(d, t-\tau)\, d\tau.$$

Since the receiver moves along the ground at a constant velocity v, the position of the receiver is $d = vt$, i.e.,

$$y(vt, t) = \int_{-\infty}^{t} x(\tau)\, h(vt, t-\tau)\, d\tau.$$

Since v is a constant, $y(vt, t)$ is just a function of t. Therefore the above equation can be expressed as

$$y(t) = \int_{-\infty}^{t} x(\tau)\, h(vt, t-\tau)\, d\tau = x(t) * h(vt, t) = x(t) * h(d, t) \tag{5.14}$$

It is useful to discretize the multipath delay axis τ of the impulse response into equal time delay segments called excess delay bins, each bin having a time delay width

equal to $(\tau_{i+1} - \tau_i) = \Delta\tau$ and $\tau_i = i\Delta\tau$ for $i \in \{0, 1, 2, ..N-1\}$, where N represents the total number of possible equally-spaced multipath components, including the first arriving component. The useful frequency span of the model is $2/\Delta\tau$. The model may be used to analyze transmitted RF signals having bandwidth less than $2/\Delta\tau$.

If there are N multipaths, maximum excess delay is given by $N\Delta\tau$.

$$\{y(t) = x(t) * h(t, \tau_i) | i = 0, 1, ...N-1\} \tag{5.15}$$

Bandpass channel impulse response model is

$$x(t) \rightarrow h(t, \tau) = Re\{h_b(t, \tau)e^{j\omega_c t}\} \rightarrow y(t) = Re\{r(t)e^{j\omega_c t}\} \tag{5.16}$$

Baseband equivalent channel impulse response model is given by

$$c(t) \rightarrow \frac{1}{2} h_b(t, \tau) \rightarrow r(t) = c(t) * \frac{1}{2} h_b(t, \tau) \tag{5.17}$$

Average power is

$$\overline{x^2(t)} = \frac{1}{2}|c(t)|^2 \tag{5.18}$$

The baseband impulse response of a multipath channel can be expressed as

$$h_b(t, \tau) = \sum_{i=0}^{N-1} a_i(t, \tau) \exp[j(2\pi f_c \tau_i(t) + \phi_i(t, \tau))]\delta(\tau - \tau_i(t)) \tag{5.19}$$

$i=0$

where $a_i(t, \tau)$ and $\tau_i(t)$ are the real amplitudes and excess delays, respectively, of the ith multipath component at time t. The phase term $2\pi f_c \tau_i(t) + \phi_i(t, \tau)$ in the above equation represents the phase shift due to free space propagation of the ith multipath component, plus any additional phase shifts which are encountered in the channel.

If the channel impulse response is wide sense stationary over a small-scale time or distance interval, then

$$h_b(\tau) = \sum_{i=0}^{N-1} a_i \exp[j\theta_i]\delta(\tau - \tau_i) \tag{5.20}$$

For measuring $h_b(\tau)$, we use a probing pulse to approximate $\delta(t)$ i.e.,

$$p(t) \approx \delta(t - \tau) \tag{5.21}$$

Power delay profile is taken by spatial average of $|h_b(t, \tau)^2|$ over a local area. The received power delay profile in a local area is given by

$$p(\tau) \approx k\overline{|h_b(t;\tau)|^2}. \tag{5.22}$$

Relation Between Bandwidth and Received Power

In actual wireless communications, impulse response of a multipath channel is measured using channel sounding techniques. Let us consider two extreme channel sounding cases.

Consider a pulsed, transmitted RF signal

$$x(t) = \mathrm{Re}\{p(t)e^{j2\pi f_c t}\} \tag{5.23}$$

where $p(t) = \sqrt{\frac{4\tau_{max}}{T_{bb}}}$ for $0 \le t \le T_{bb}$ and 0 elsewhere. The low pass channel output is

$$\begin{aligned} r(t) &= \frac{1}{2}\sum_{i=0}^{N-1} a_i \exp[j\theta_i]p(t - \tau_i) \\ &= \sum_{i=0}^{N-1} a_i \exp[j\theta_i] \cdot \sqrt{\frac{\tau_{max}}{T_{bb}}}\, \mathrm{rect}\left(t - \frac{T_b}{2} - \tau_i\right). \end{aligned}$$

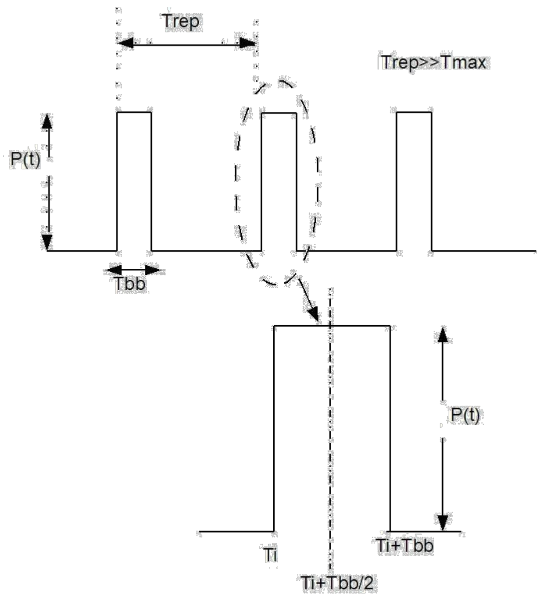


Figure 5.2: A generic transmitted pulsed RF signal.

The received power at any time t_0 is

$$|r(t_0)|^2 = \frac{1}{\tau_{max}} \int_0^{\tau_{max}} r(t) r^*(t) dt$$

$$= \frac{1}{\tau_{max}} \int_0^{\tau_{max}} \frac{1}{4} \left(\sum_{k=0}^{N-1} a_k^2(t_0) p^2(t - \tau_k) \right) dt$$

$$= \frac{1}{\tau_{max}} \sum_{k=0}^{N-1} a_k^2(t_0) \int_0^{\tau_{max}} \left\{ \sqrt{\frac{\tau_{max}}{T_{bb}}} \, rect\left(t - \frac{T_{bb}}{2} - \tau_i\right) \right\}^2 dt$$

$$= \sum_{k=0}^{N-1} a_k^2(t_0).$$

Interpretation: If the transmitted signal is able to resolve the multipaths, then average small-scale receiver power is simply sum of average powers received from each multipath components.

$$E_{a,\theta}[P_{WB}] = E_{a,\theta}\left[\sum_{i=0}^{N-1} |a_i \exp(j\theta_i)|^2\right] \approx \sum_{i=0}^{N-1} \overline{a_i^2} \tag{5.24}$$

Now instead of a pulse, consider a CW signal, transmitted into the same channel and for simplicity, let the envelope be c(t) = 2. Then

$$r(t) = \sum_{i=0}^{N-1} a_i \exp[j\theta_i(t, \tau)] \tag{5.25}$$

and the instantaneous power is

$$|r(t)|^2 = \left|\sum_{i=0}^{N-1} a_i \exp[j\theta_i(t,\tau)]\right|^2 \quad (5.26)$$

Over local areas, a_i varies little but θ_i varies greatly resulting in large fluctuations.

$$E_{a,\theta}[P_{CW}] = E_{a,\theta}\left[\sum_{i=0}^{N-1} |a_i \exp(j\theta_i)|^2\right]$$

$$\approx \sum_{i=0}^{N-1} \overline{a_i^2} + 2\sum_{i=0}^{N-1}\sum_{i,j\neq i}^{N} r_{ij}\,\overline{\cos(\theta_i - \theta_j)}$$

where $r_{ij} = E_a[a_i a_j]$.

If, $r_{ij} = \overline{\cos(\theta_i - \theta_j)} = 0$, then $E_{a,\theta}[P_{CW}] = E_{a,\theta}[P_{WB}]$. This occurs if multipath components are uncorrelated or if multipath phases are i.i.d over $[0, 2\pi]$.

Bottomline:

1. If the signal bandwidth is greater than multipath channel bandwidth then fading effects are negligible

2. If the signal bandwidth is less than the multipath channel bandwidth, large fading occurs due to phase shift of unresolved paths.

Linear Time Varying Channels (LTV)

The time variant transfer function(TF) of an LTV channel is FT of $h(t, \tau)$ w.r.t. τ.

$$H(f,t) = FT[h(\tau,t)] = \int_{-\infty}^{\infty} h(\tau,t)e^{-j2\pi f\tau}\,d\tau \quad (5.27)$$

$$h(\tau,t) = FT^{-1}[H(f,t)] = \int_{-\infty}^{\infty} H(f,t)e^{j2\pi f\tau}\,df \quad (5.28)$$

The received signal

$$r(t) = \int_{-\infty}^{\infty} R(f,t)e^{j2\pi ft}\,df \quad (5.29)$$

where $R(f,t) = H(f,t)X(f)$.

For flat fading channel, $h(\tau,t) = Z(t)\delta(\tau - \tau_i)$ where $Z(t) = \sum \alpha_n(t)e^{-j2\pi f_c\tau_n(t)}$. In this case, the received signal is

$$r(t) = \int_{-\infty}^{\infty} h(\tau,t)x(t-\tau)\,d\tau = Z(t)x(t-\tau_i) \quad (5.30)$$

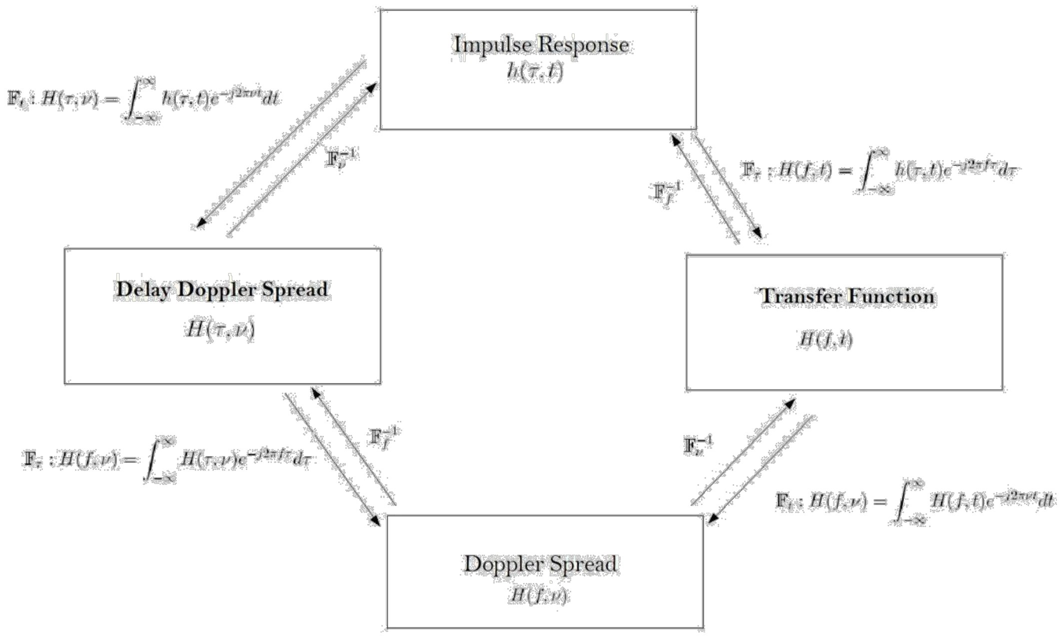


Figure 5.3: Relationship among different channel functions.

where the channel becomes multiplicative.

Doppler spread functions:

$$H(f, \nu) = FT\,[H(f, t)] = \int_{-\infty}^{\infty} H(f, t)e^{-j2\pi\nu t}dt \qquad (5.31)$$

and

$$H(f, t) = FT^{-1}[H(f, \nu)] = \int_{-\infty}^{\infty} H(f, \nu)e^{j2\pi\nu t}d\nu \qquad (5.32)$$

Delay Doppler spread:

$$H(\tau, \nu) = FT\,[h(\tau, t)] = \int_{-\infty}^{\infty} h(\tau, t)e^{-j2\pi\nu t}dt \qquad (5.33)$$

Small-Scale Multipath Measurements

Direct RF Pulse System:

A wideband pulsed bistatic radar usually transmits a repetitive pulse of width T_{bb} s, and uses a receiver with a wide bandpass filter ($BW = \frac{2}{T_{bb}}$ Hz). The signal is then amplified, envelope detected, and displayed and stored on a high speed oscilloscope. Immediate measurements of the square of the channel impulse response convolved with the probing pulse can be taken. If the oscilloscope is set on averaging mode, then this system provides a local average power delay profile.

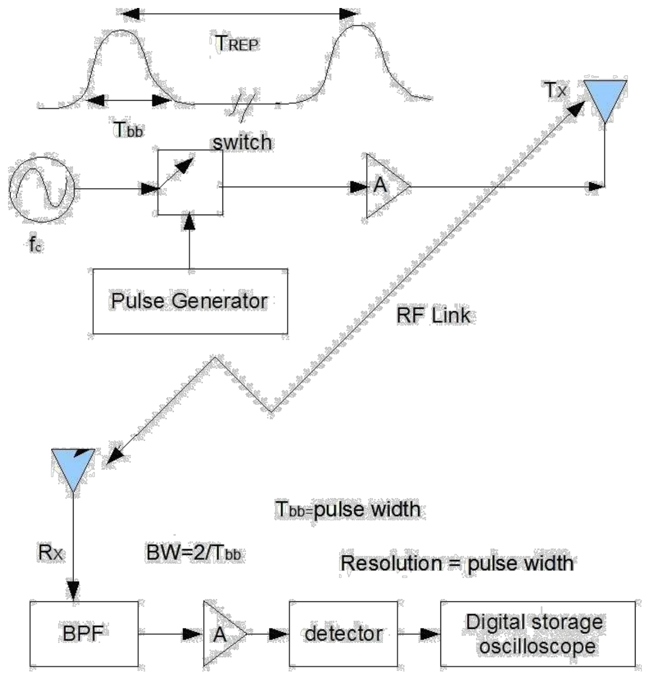


Figure 5.4: Direct RF pulsed channel IR measurement.

This system is subject to interference noise. If the first arriving signal is blocked or fades, severe fading occurs, and it is possible the system may not trigger properly.

Frequency Domain Channel Sounding

In this case we measure the channel in the frequency domain and then convert it into time domain impulse response by taking its inverse discrete Fourier transform (IDFT). A vector network analyzer controls a swept frequency synthesizer. An S-parameter test set is used to monitor the frequency response of the channel. The sweeper scans a particular frequency band, centered on the carrier, by stepping through discrete frequencies. The number and spacing of the frequency step impacts the time resolution of the impulse response measurement. For each frequency step, the S-parameter test set transmits a known signal level at port 1 and monitors the received signal at port 2. These signals allow the analyzer to measure the complex response, $S_{21}(\omega)$, of the channel over the measured frequency range. The $S_{21}(\omega)$ measure is the measure of the signal flow from transmitter antenna to receiver

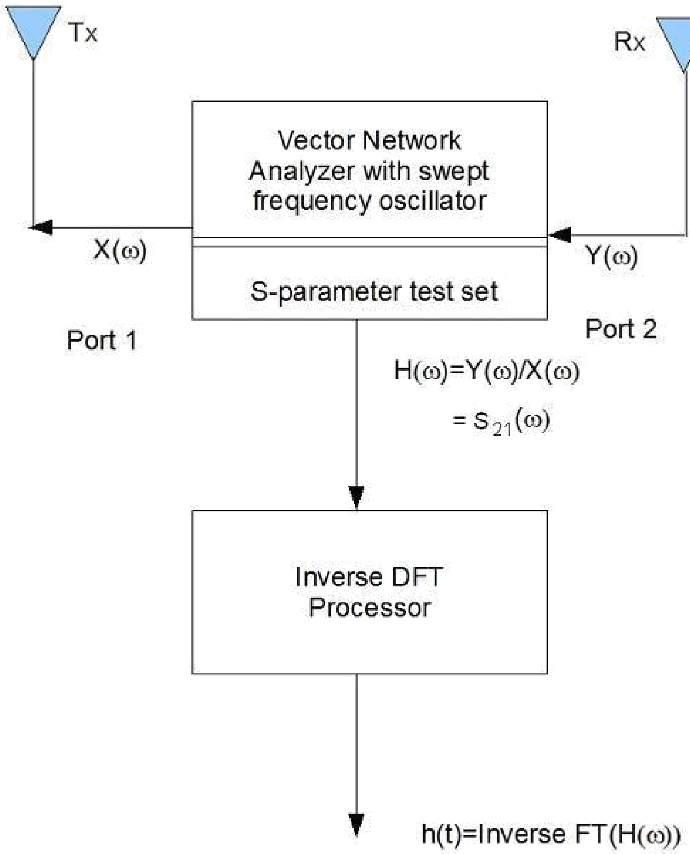


Figure 5.5: Frequency domain channel IR measurement.

antenna (i.e., the channel).

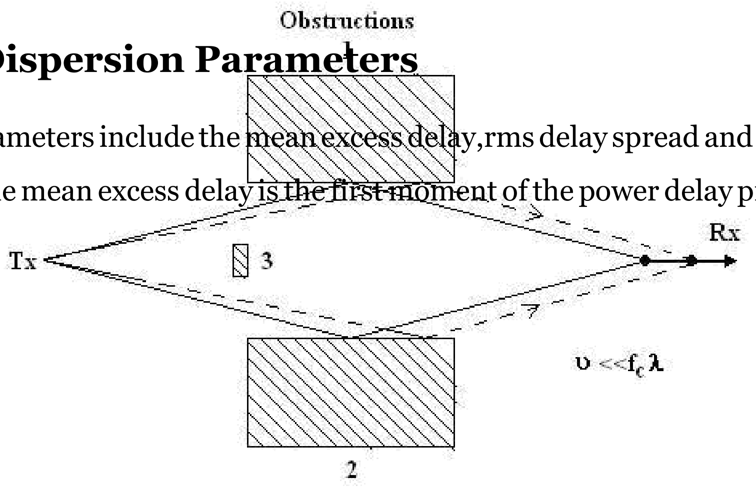
This system is suitable only for indoor channel measurements. This system is also non real-time. Hence, it is not suitable for time-varying channels unless the sweep times are fast enough.

Parameters of Mobile Multipath Channel

To compare the different multipath channels and to quantify them, we define some parameters. They all can be determined from the power delay profile. These parameters can be broadly divided in to two types.

Time Dispersion Parameters

These parameters include the mean excess delay,rms delay spread and excess delay spread. The mean excess delay is the first moment of the power delay profile and is



defined as

$$\bar{\tau} = \frac{\sum_k a_k^2 \tau_k}{\sum_k a_k^2} = \frac{\sum_k P(\tau_k)\tau_k}{\sum_k P(\tau_k)} \qquad (5.34)$$

where a_k is the amplitude, τ_k is the excess delay and $P(\tau_k)$ is the power of the individual multipath signals.

The mean square excess delay spread is defined as

$$\overline{\tau^2} = \frac{\sum_k P(\tau_k)\tau_k^2}{\sum_k P(\tau_k)} \qquad (5.35)$$

Since the rms delay spread is the square root of the second central moment of the power delay profile, it can be written as

$$\sigma_\tau = \sqrt{\overline{\tau^2} - (\bar{\tau})^2} \qquad (5.36)$$

As a rule of thumb, for a channel to be flat fading the following condition must be satisfied

$$\frac{\sigma_\tau}{T_S} \leq 0.1 \qquad (5.37)$$

where T_S is the symbol duration. For this case, no equalizer is required at the receiver.

Example 2

1. Sketch the power delay profile and compute RMS delay spread for the following:

$$P(\tau) = \sum_{n=0}^{1} \delta(\tau - n \times 10^{-6}) \text{ (in watts)}$$

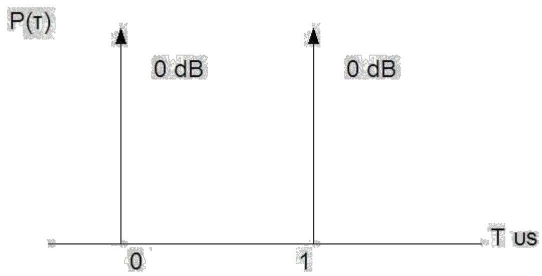
2. If BPSK modulation is used, what is the maximum bit rate that can be sent through the channel without needing an equalizer?

Solution

1. $P(0) = 1$ watt, $P(1) = 1$ watt

$$\bar{\tau} = \frac{(1)(0) + (1)(1)}{1 + 1} = 0.5\mu s$$

$$\overline{\tau^2} = 0.5\mu s^2 \quad \sigma_\tau = 0.5\mu s$$

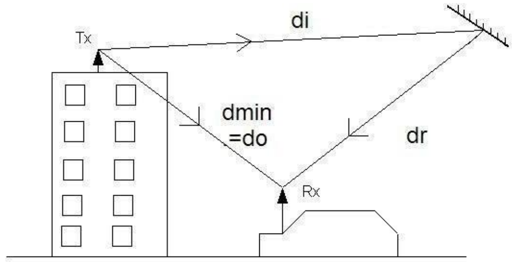


2. For flat fading channel, we need $\frac{\sigma_\tau}{T_s} 0.1 \Rightarrow R_s = \frac{1}{T_s} = 0.2 \times 10^4 = 200$ kbps

For BPSK we need $R_b = R_s = 200$ kbps

Example 3 A simple delay spread bound: **Feher's upper bound**

Consider a simple worst-case delay spread scenario as shown in figure below.



Here $d_{min} = d_0$ and $d_{max} = d_i + d_r$

Transmitted power = P_T, Minimum received power = $P_{R_{min}} = P_{Threshold}$

$$\frac{P_{R_{min}}}{P_T} = tt_T \, tt_R \left(\frac{\lambda}{4\pi d_{max}}\right)^2$$

Put $tt_T = tt_R = 1$ i.e., considering omni-directional unity gain antennas

$$d_{max} = \left(\frac{\lambda}{4\pi}\right)\left(\frac{P_T}{P_{R_{min}}}\right)^{\frac{1}{2}}$$

$$\tau_{max} = \frac{d_{max}}{c} = \left(\frac{\lambda}{4\pi c}\right)\left(\frac{P_T}{P_{R_{min}}}\right)^{\frac{1}{2}}$$

$$\tau_{max} = \left(\frac{1}{4\pi f}\right)\left(\frac{P_T}{P_{R_{min}}}\right)^{\frac{1}{2}}$$

Frequency Dispersion Parameters

To characterize the channel in the frequency domain, we have the following param-eters.

Coherence bandwidth:

It is a statistical measure of the range of frequencies over which the channel can be considered to pass all the frequency components with almost equal gain and linear phase. When this condition is satisfied then we say the channel to be flat.

Practically, coherence bandwidth is the minimum separation over which the two frequency components are affected differently. If the coherence bandwidth is con-sidered to be the bandwidth over which the frequency correlation function is above 0.9, then it is approximated as

$$B_C \approx \frac{1}{50\sigma_\tau} . \qquad (5.38)$$

However, if the coherence bandwidth is considered to be the bandwidth over which the frequency correlation function is above 0.5, then it is defined as

$$B_C \approx \frac{1}{5\sigma_\tau} . \qquad (5.39)$$

The coherence bandwidth describes the time dispersive nature of the channel in the local area. A more convenient parameter to study the time variation of the channel is the coherence time. This variation may be due to the relative motion between the mobile and the base station or the motion of the objects in the channel.

Coherence time:

This is a statistical measure of the time duration over which the channel impulse response is almost invariant. When channel behaves like this, it is said to be slow faded. Essentially it is the minimum time duration over which two received signals are affected differently. For an example, if the coherence time is considered to be the bandwidth over which the time correlation is above 0.5, then it can be approximated as

$$T_C \approx \frac{9}{16\pi f_m} \qquad (5.40)$$

where f_m is the maximum doppler spread given be $f_m = \frac{v}{\lambda}$.

Doppler spread:

Another parameter is the Doppler spread (B_D) which is the range of frequencies over which the received Doppler spectrum is non zero.

Types of Small-Scale Fading

The type of fading experienced by the signal through a mobile channel depends on the relation between the signal parameters (bandwidth, symbol period) and the channel parameters (rms delay spread and Doppler spread). Hence we have four different types of fading. There are two types of fading due to the time dispersive nature of the channel.

Fading Effects due to Multipath Time Delay Spread

Flat Fading

Such types of fading occurs when the bandwidth of the transmitted signal is less than the coherence bandwidth of the channel. Equivalently if the symbol period of the signal is more than the rms delay spread of the channel, then the fading is flat fading.

So we can say that flat fading occurs when

$$B_S \quad B_C \tag{5.1}$$

where B_S is the signal bandwidth and B_C is the coherence bandwidth. Also

$$T_S \quad \Sigma_T \tag{5.2}$$

where T_S is the symbol period and σ_T is the rms delay spread. And in such a case, mobile channel has a constant gain and linear phase response over its bandwidth.

Frequency Selective Fading

Frequency selective fading occurs when the signal bandwidth is more than the coherence bandwidth of the mobile radio channel or equivalently the symbols duration of the signal is less than the rms delay spread.

$$B_S \quad B_C \tag{5.3}$$

and

$$T_S \quad \Sigma_T \tag{5.4}$$

At the receiver, we obtain multiple copies of the transmitted signal, all attenuated and delayed in time. The channel introduces inter symbol interference. A rule of thumb for a channel to have flat fading is if

$$\frac{\Sigma_T}{T_S} \leq 0.1 \qquad (5.5)$$

Fading Effects due to Doppler Spread

Fast Fading

In a fast fading channel, the channel impulse response changes rapidly within the symbol duration of the signal. Due to Doppler spreading, signal undergoes frequency dispersion leading to distortion. Therefore a signal undergoes fast fading if

$$T_s \quad T_C \qquad (5.6)$$

where T_C is the coherence time and

$$B_S \quad B_D \qquad (5.7)$$

where B_D is the Doppler spread. Transmission involving very low data rates suffer from fast fading.

Slow Fading

In such a channel, the rate of the change of the channel impulse response is much less than the transmitted signal. We can consider a slow faded channel a channel in which channel is almost constant over atleast one symbol duration. Hence

$$T_s \quad T_C \qquad (5.8)$$

and

$$B_S \quad B_D \qquad (5.9)$$

We observe that the velocity of the user plays an important role in deciding whether the signal experiences fast or slow fading.

Statistical models for multipath propagation

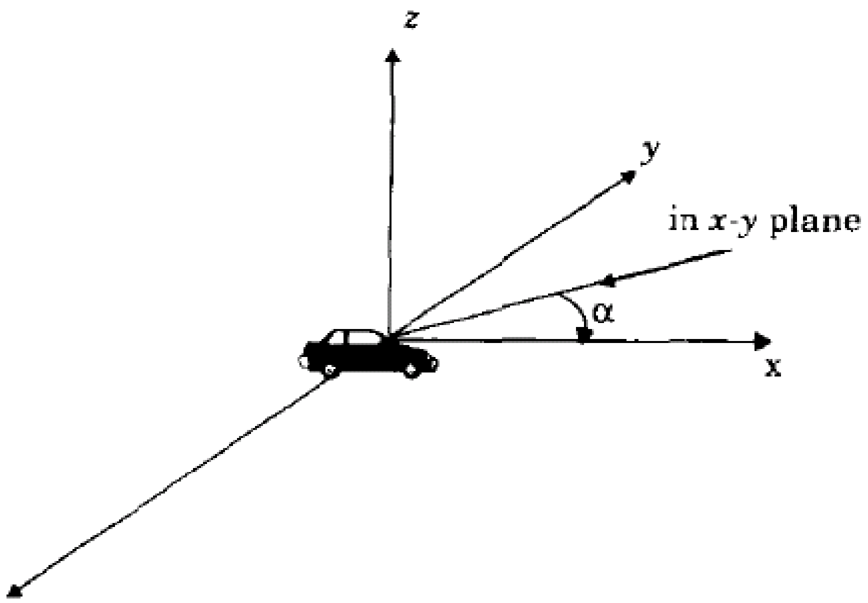
4.7.1 Clarke's Model for Flat Fading

Clarke [Cla68] developed a model where the statistical characteristics of the electromagnetic fields of the received signal at the mobile are deduced from scattering. The model assumes a fixed transmitter with a vertically polarized antenna. The field incident on the mobile antenna is assumed to be comprised of N azimuthal plane waves with arbitrary carrier phases, arbitrary azimuthal angles of arrival, and each wave having equal average amplitude. It should be noted that the equal average amplitude assumption is based on the fact that in the absence of a direct line-of-sight path, the scattered components arriving at a receiver will experience similar attenuation over small-scale distances.

Figure 4.19 shows a diagram of plane waves incident on a mobile traveling at a velocity v, in the x-direction. The angle of arrival is measured in the x-y plane with respect to the direction of motion. Every wave that is incident on the mobile undergoes a Doppler shift due to the motion of the receiver and arrives at the receiver at the same time. That is, no excess delay due to multipath is assumed for any of the waves (flat fading assumption). For the n th wave arriving at an angle α_n to the x-axis, the Doppler shift in Hertz is given by

$$f_n = \frac{v}{\lambda}\cos\alpha_n$$

where λ is the wavelength of the incident wave.



Illustrating plane waves arriving at random angles.

The vertically polarized plane waves arriving at the mobile have E and H field components given by

$$E_z = E_o \sum^{N} C_n \cos(2\pi f_c t + \theta_n)$$

$$H_y = \frac{``a}{\eta} \sum C_n \cos\alpha_n \cos(2\pi f_c t + \theta_n)$$

where *o ia the real amplitude of lotal average E-field 1&BISl2fRed constants, Cq is a real rsndom variable representing the amplitude of individual waves, r| is the intrinsic impedance of free space i 37711), and /, is the carrier frequency. the random phaae of the n th arriving component 8 is given by

8 2 ,5 ' + *In*

The amplitudes of the E-and H-field are normalized such that the ensemble average of the Cg 's ia given by

Since the Doppler shift is very small when compared to the carrier frequency, the three field components may be modeled as narrow band random processes. The three componentB *E . Hq,* and Hg can be approximated as Gaussian random variables if A' is sufficiently large. The phase angles are assumed to have a uniform probability density function (pdf) on the interval (O,2nl. Based on the analysis by Rice (Ric481 the E-field can be expressed in an in-phase and quadrature form

wwhere :2 c

*Bo*th *T I I)* and P, (f) are Gaussian random proeeaeea which are denoted aa T, and T', , respectively, at any line t. T, and T, are uncorrelated zero-mean Gaussian random variables with an equal variance given by

$$\overline{T_c^2} = \overline{T_s^2} \quad \overline{|E_z|^2} \quad E_0^2/2$$

mheretheoverbsrdenotestheeneembleaveragm

RemveJopeofthererei v m-Geld,F f,*s venby

*J!!, l*1| -‘ •1Z*c [t) +* , (1 r(‹)

Since T, and fi, are Gauaaian random variables, it can be shown through a Jacobean transformation [Pap91J that the random received signal envelope r has a Rayleigh distribution given by

$$p(r) \begin{cases} \frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right) & 0 \le r \le \infty \\ 0 & r < 0 \end{cases}$$

wbere a' = Z' /2

4.7.1.1 Spectral Shape Due to Dopplw Sp›eed In Clarke'c Model

Gans (Gan72) developed a spectrum analysis for Clarke's model. Let *p l a da* denote the fraction of the total incomin g power within *da* of the angle o, and let A denote the average served power with reapert to ari isotropic antenna. As N —+ ‹o . *p I a) da* approaches a continuous, rather than a discrete, distribution. If G (‹i) is the azimuthal gain **pattern of** the mobile antenna as a function **of the angle of** arrival, the total received power **can be expressed as**

where *AT (IR I O)da* is the differential variation of received power with angle. If the scattered signal is a CW signal of frequency /, , then the instantaneous frequency o(the received eignal component arriving at an angle o is obtained using equation (4.Str

$$f(\alpha) \quad f \quad \frac{v}{\lambda}\cos(\alpha) + f_c \quad f_m \cos\alpha + f_c$$

where Jp ia the ttiaximum Doppler ahifi. It should be noted that /(n) ia an even function of o, (i.e., /fcd = J—ct)l.

If 6 (/) is the power spectrum of the received signal, the differential variation **ofreceived** wer with frequency ia given bv

$$S(f)|df|$$

Equating the différential variation of received power with frequency to the differential variation in received power with angle, we have

S {/} |*df*| -- *A @)o) C (a)* +p (—<t) G (—a)] |rf o I

differentiating equation (4.TO), and rearranging the terms, we have

$$|df| \quad |d\alpha||-\sin\alpha|f_m$$

Using equation (4.701, n can be expressed as a function of aa

This implies that

$$\sqrt{1-\left(\frac{f-f_c}{f_m}\right)^2}$$

i3ubstituting equation (4.73) and (d.75) into both aides of (4.72), the power spectral density *S* (,I) can be exqreased as

$$S(f) = \frac{A\left[p(\alpha)G(\alpha)+p(-\alpha)G(-\alpha)\right]}{f_m\sqrt{1-\left(\frac{f-f_c}{f_m}\right)^2}}$$

where

$$S(f) \quad 0, \qquad |f-f_c| > f_m$$

The spectrum is centered on the her frequency and iB zero outside the limits of/ z „,,. Each of the arriving waves has ita own carrier frequency (due to ito direction of arrival) which is slightly offset from the center frequency. For the caae of a vertical XZ4 antenna (fi (o) = 1.5), and a uniform distribution jo(a) = **I<2n over O to 2n,theoutpmLapectruoziegivsnby**

In equation (4.78) the power spectral density at = J, z/q ia infinite, i.e., Doppler components arriving at exactly 0° and 180° have an infinite power spectral density. This i8 not a problem since n is continuously distributed and the probahility of components arriving at exactly theae anglea is zero.

Figure 4.20 showa the power spectral density of the resulting RF signal due to Doppler fading. Smith fSmi75) demonstrated an easy way to simulate Clarke's model using a computer simulation ma described Section 4.7.2.

After envelope detection of the Doppler-shifted aignal, the resulting baseband spectrum he a maximum frequency of 2Jq . It can be shown fJak7<) that the electric field produces a baaeband power spectral density given by

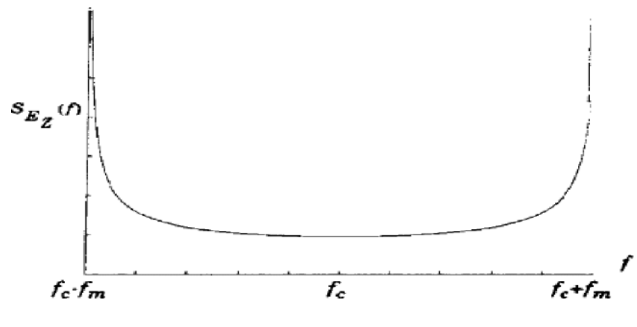


Figure 4.20

$$S_{bbE_z}(f) \quad \frac{1}{8\pi f_m}K\left[\sqrt{1-\left(\frac{f}{2f_m}\right)^2}\right]$$

where A (•) is the complete elliptical integral of the firat kind. Equation (4.79) ia not **intuitive** and is a result of the temporal correlation of the received signal wheat passed through a nonlinear **envelope detector. Figure** 4.21 illustrated the baaeb nd spectrum of the received signal after envelope detection.

The spectral shape of the Doppler spread determines the time domain fading waveform and dictates the temporal correlation and fade slope behaviors. Rayleigh fading simulators must use a fading spectrum such as equation (4.78) in order to produce realistic fading waveforms that have proper time correlation.

Many multipath models have been proposed to explain the observed statistical nature of a practical mobile channel. Both the first order and second order statistics

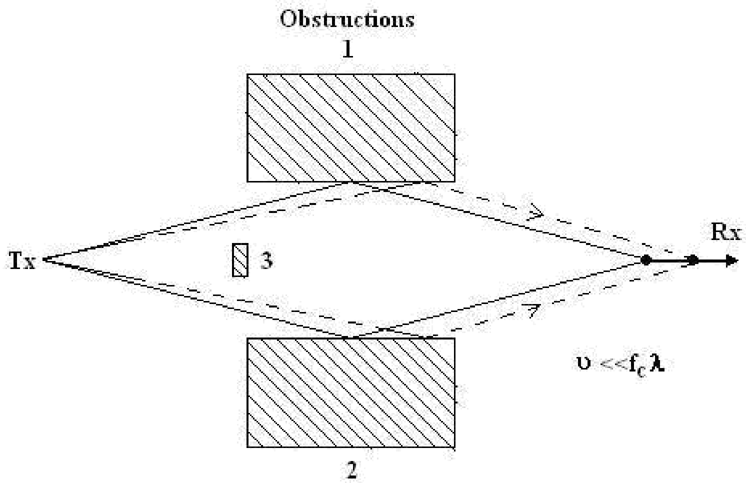


Figure 5.6: Two ray NLoS multipath, resulting in Rayleigh fading.

have been examined in order to find out the effective way to model and combat the channel effects. The most popular of these models are Rayleigh model, which de-scribes the NLoS propagation. The Rayleigh model is used to model the statistical time varying nature of the received envelope of a flat fading envelope. Below, we discuss about the main first order and second order statistical models.

NLoS Propagation: Rayleigh Fading Model

Let there be two multipath signals S_1 and S_2 received at two different time instants due to the presence of obstacles as shown in Figure 5.6. Now there can either be constructive or destructive interference between the two signals.

Let E_n be the electric field and Θ_n be the relative phase of the various multipath signals.So we have

$$E = \sum_{n=1}^{N} E_n e^{j\theta_n}$$

Now if $N \to \infty$(i.e. are sufficiently large number of multipaths) and all the E_n are IID distributed, then by Central Limit Theorem we have,

$$\lim_{N \to \infty} \tilde{E} = \lim_{N \to \infty} \sum_{n=1}^{N} E_n e^{j\theta_n} \quad (5.42)$$

$$= Z_r + jZ_i = Re^{j\varphi} \quad (5.43)$$

where Z_r and Z_i are Gaussian Random variables. For the above case

$$R = \sqrt{Z_r^2 + Z_i^2} \tag{5.44}$$

and

$$\varphi = \tan^{-1}\frac{Z_i}{Z_r} \tag{5.45}$$

For all practical purposes we assume that the relative phase Θ_n is uniformaly distributed.

$$E[e^{j\theta_n}] = \frac{1}{2\pi}\int_0^{2\pi} e^{j\theta}\, d\theta = 0 \tag{5.46}$$

It can be seen that E_n and Θ_n are independent. So,

$$E[\tilde{E}] = E[\sum E_n e^{j\theta_n}] = 0 \tag{5.47}$$

$$E[|\tilde{E}|^2] = E[\sum_n E_n e^{j\theta_n} \sum_n E_n^* e^{-j\theta_n}] = E[\sum_m \sum_n E_m E_n e^{j(\theta_n - \theta_m)}] = \sum_{n=1}^{N} E_n^2 = P_0 \tag{5.48}$$

where P_0 is the total power obtained. To find the Cumulative Distribution Function(CDF) of R, we proceed as follows.

$$F_R(r) = P_r(R \le r) = \iint_A f_{Z_i,Z_r}(z_i, z_r)\, dz_i dz_r \tag{5.49}$$

where A is determined by the values taken by the dummy variable r. Let Z_i and Z_r be zero mean Gaussian RVs. Hence the CDF can be written as

$$F_R(r) = \iint_A \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(Z_r^2+Z_i^2)}{2\sigma^2}} dZ_i dZ_r \tag{5.50}$$

Let $Z_r = p\cos(\Theta)$ and $Z_i = p\sin(\Theta)$ So we have

$$F_R(r) = \int_0^{2\pi}\int_0^{2\pi} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-p^2}{2\sigma^2}} p\,dp\,d\theta \tag{5.51}$$

$$= 1 - e^{\frac{-r^2}{2\sigma^2}} \tag{5.52}$$

Above equation is valid for all $r \ge 0$. The pdf can be written as

$$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \tag{5.53}$$

and is shown in Figure 5.7 with different σ values. This equation too is valid for all $r \ge 0$. Above distribution is known as Rayleigh distribution and it has been derived

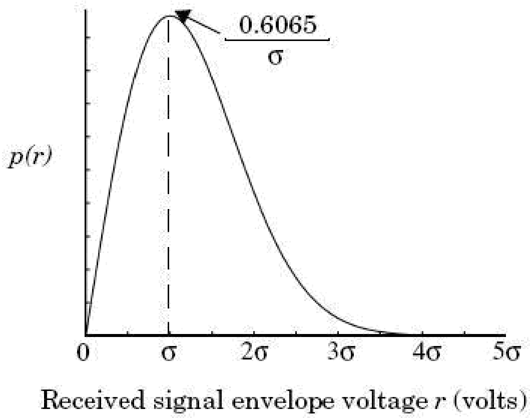


Figure 5.7: Rayleigh probability density function.

for slow fading. However, if f_D 1 Hz, we call it as Quasi-stationary Rayleigh fading. We observe the following:

$$E[R] = \sqrt{\frac{\pi}{2}}\sigma \tag{5.54}$$

$$E[R^2] = 2\sigma^2 \tag{5.55}$$

$$var[R] = (2 - \frac{\pi}{2})\sigma^2 \tag{5.56}$$

$$median[R] = 1.77\sigma. \tag{5.57}$$

LoS Propagation: Rician Fading Model

Rician Fading is the addition to all the normal multipaths a direct LOS path.

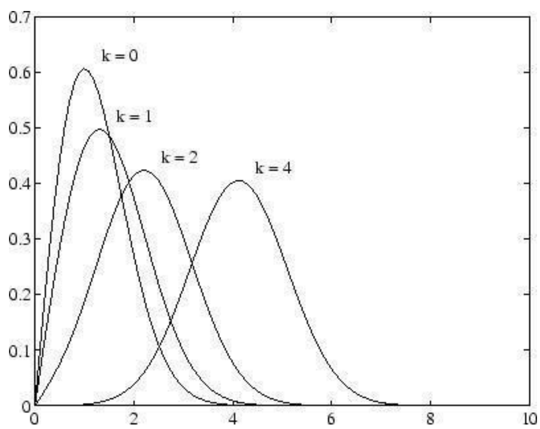


Figure 5.8: Ricean probability density function.

$$f_R(r) = \frac{r}{\sigma^2} e^{\frac{-(r^2+A^2)}{2\sigma^2}} I_0\left(\frac{Ar}{\sigma^2}\right) \tag{5.58}$$

for all A ≥ 0 and r ≥ 0. Here A is the peak amplitude of the dominant signal and $I_0(.)$ is the modified Bessel function of the first kind and zeroth order.

A factor K is defined as

$$K_{dB} = 10 \log \frac{A^2}{2\sigma^2} \tag{5.59}$$

As $A \to 0$ then $K_{dB} \to \infty$.

Generalized Model: Nakagami Distribution

A generalization of the Rayleigh and Rician fading is the Nakagami distribution.

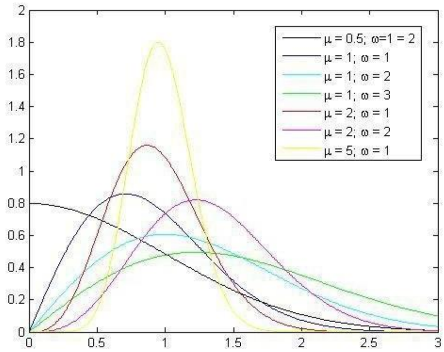


Figure 5.9: Nakagami probability density function.

Its pdf is given as,

$$f_R(r) = \frac{2r^{m-1}}{\Gamma(m)\,\Omega^m}\left(\frac{m}{m}\right)e^{\frac{-mr^2}{\Omega}} \tag{5.60}$$

where,

$\Gamma(m)$ is the gamma function

Ω is the average signal power and

m is the fading factor. It is always greater than or equal to 0.5.

When m=1, Nakagami model is the Rayleigh model.

When

$$m = \frac{(M+1)^2}{2M+1}$$

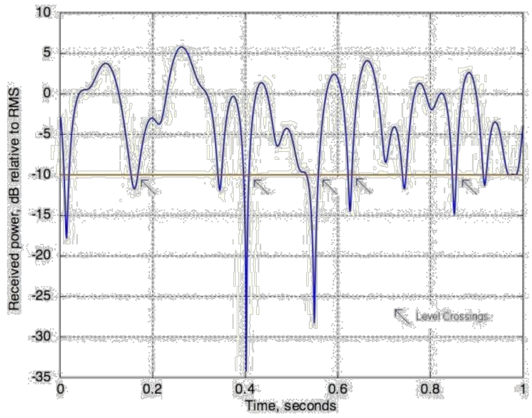


Figure 5.10: Schematic representation of level crossing with a Rayleigh fading envelope at 10 Hz Doppler spread.

where

$$M = \frac{A}{2\sigma}$$

Nakagami fading is the Rician fading.
As $m \to \infty$ Nakagami fading is the impulse channel and no fading occurs.

Second Order Statistics

To design better error control codes, we have two important second order parameters of fading model, namely the **level crossing rate** (LCR) and **average fade duration** (AFD). These parameters can be utilized to assess the speed of the user by measuring them through the reverse channel. The LCR is the expected rate at which the Rayleigh fading envelope normalized to the local rms amplitude crosses a specific level 'R' in a positive going direction.

$$N_R = \int_0^\infty \dot{r}\, p(R, \dot{r})\, d\dot{r} = \sqrt{2\pi} f_D \rho e^{-\rho^2} \tag{5.61}$$

where $\dot{r}$ is the time derivative of r(t), f_D is the maximum Doppler shift and ρ is the value of the specified level R, normalized to the local rms amplitude of the fading envelope.

The other important parameter, AFD, is the average period time for which the receiver power is below a specified level R.

$$\bar{\tau} = \frac{1}{N_r} P_r(r \le R) \tag{5.62}$$

As

$$P_r(r \le R) = \int_0^R p(r)dr = 1 - e^{-\rho^2}, \tag{5.63}$$

therefore,

$$\bar{\tau} = \frac{1 - e^{-\rho^2}}{\sqrt{2\pi} f_D \rho e^{-\rho^2}} \tag{5.64}$$

$$= \frac{e^{\rho^2} - 1}{\sqrt{2\pi} f_D \rho}. \tag{5.65}$$

Apart from LCR, another parameter is fading rate, which is defined as the number of times the signal envelope crosses the middle value (r_m) in a positive going direction per unit time. The average rate is expressed as

$$N(r_m) = \frac{2v}{\lambda}. \tag{5.66}$$

Another statistical parameter, sometimes used in the mobile communication, is called as depth of fading. It is defined as the ratio between the minimum value and the mean square value of the faded signal. Usually, an average value of 10% as depth of fading gives a marginal fading scenario.

Simulation of Rayleigh Fading Models

Clarke's Model: without Doppler Effect

In it, two independent Gaussian low pass noise sources are used to produce in-phase and quadrature fading branches. This is the basic model and is useful for slow fading channel. Also the Doppler effect is not accounted for.

Clarke and Gans' Model: with Doppler Effect

In this model, the output of the Clarke's model is passed through Doppler filter in the RF or through two initial baseband Doppler filters for baseband processing as shown in Figure 5.11. Here, the obtained Rayleigh output is flat faded signal but not frequency selective.

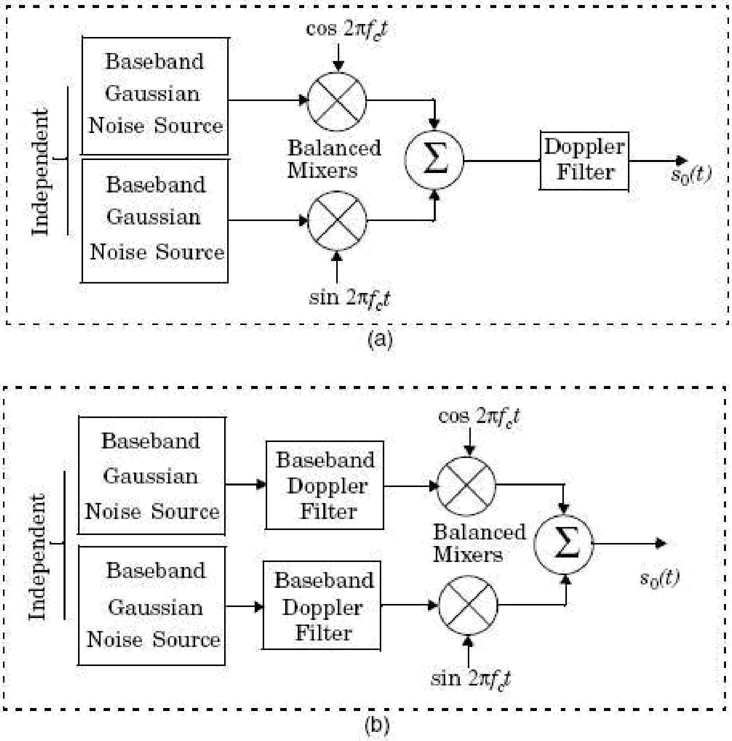


Figure 5.11: Clarke and Gan's model for Rayleigh fading generation using quadrature amplitude modulation with (a) RF Doppler filter, and, (b) baseband Doppler filter.

Rayleigh Simulator with Wide Range of Channel Conditions

To get a frequency selective output we have the following simulator through which both the frequency selective and flat faded Rayleigh signal may be obtained. This is achieved through varying the parameters a_i and τ_i, as given in Figure 5.12.

Two-Ray Rayleigh Faded Model

The above model is, however, very complex and difficult to implement. So, we have the two ray Rayleigh fading model which can be easily implemented in software as shown in Figure 5.13.

$$h_b(t) = \alpha_1 e^{j\varphi_1} \delta(t) + \alpha_2 e^{j\varphi_2} \delta(t - \tau) \quad (5.67)$$

where α_1 and α_2 are independent Rayleigh distributed and φ_1 and φ_2 are independent and uniformaly distributed over 0 to 2π. By varying τ it is possible to create a wide range of frequency selective fading effects.

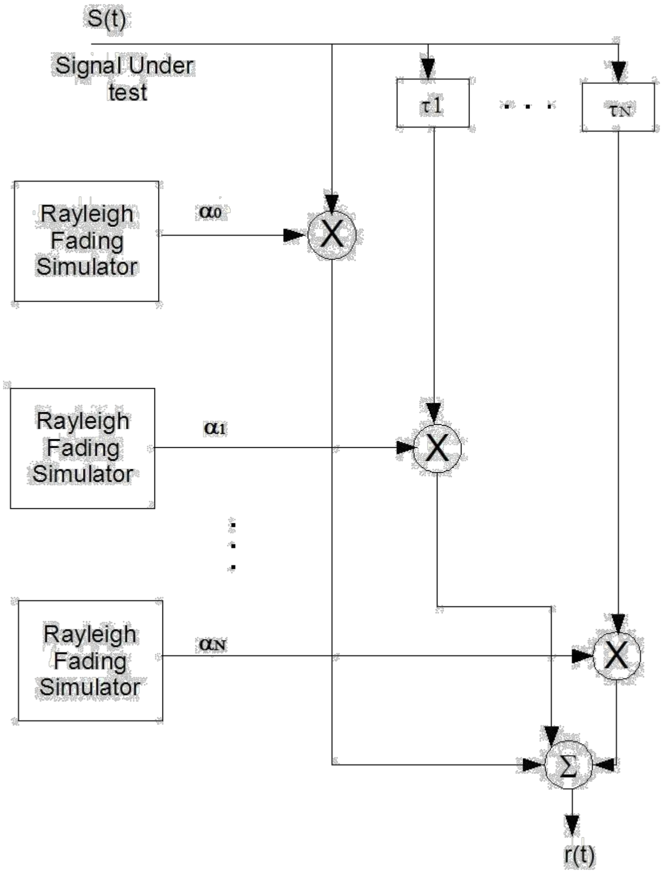


Figure 5.12: Rayleigh fading model to get both the flat and frequency selective channel conditions.

Saleh and Valenzuela Indoor Statistical Model

This method involved averaging the square law detected pulse response while sweeping the frequency of the transmitted pulse. The model assumes that the multipath components arrive in clusters. The amplitudes of the received components are independent Rayleigh random variables with variances that decay exponentially with cluster delay as well as excess delay within a cluster. The clusters and multipath components within a cluster form Poisson arrival processes with different rates.

SIRCIM/SMRCIM Indoor/Outdoor Statistical Models

SIRCIM (Simulation of Indoor Radio Channel Impulse-response Model) generates realistic samples of small-scale indoor channel impulse response measurements. Sub-

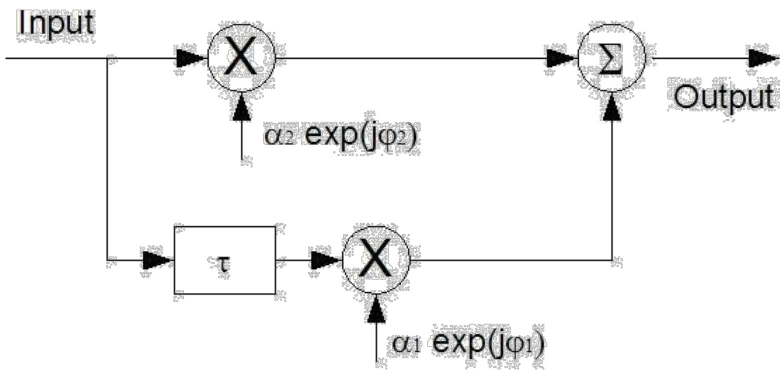


Figure 5.13: Two-ray Rayleigh fading model.

sequent work by Huang produced SMRCIM (Simulation of Mobile Radio Channel Impulse-response Model), a similar program that generates small-scale urban cellular and micro-cellular channel impulse responses.

4.7.3 Level Crossing and Fading Statistics

Rice computed joint statistics for a mathematical problem which is similar to Clarke's fading model [Cla68], and thereby provided simple expressions for computing the average number of level crossing and the duration of fades. The *level crossing rate* (LCR) and *average fade duration* of a Rayleigh fading signal are two important statistics which are useful for designing error control codes and diversity schemes to be used in mobile communication systems, since it becomes possible to relate the time rate of change of the received signal to the signal level and velocity of the mobile.

The *level crossing rate* (LCR) is defined as the expected rate at which the Rayleigh fading envelope, normalized to the local rms signal level, crosses a specified level in a positive-going direction. The number of level crossings per second is given by

$$N_R = \int_0^\infty \dot{r}\, p(R, \dot{r})\, d\dot{r} = \sqrt{2\pi} f_m \rho e^{-\rho^2} \tag{4.80}$$

where $\dot{r}$ is the time derivative of $r(t)$ (i.e., the slope), $p(R, \dot{r})$ is the joint density function of r and $\dot{r}$ at $r = R$, f_m is the maximum Doppler frequency and $\rho = R/R_{rms}$ is the value of the specified level R, normalized to the local rms amplitude of the fading envelope [Jak74]. Equation (4.80) gives the value of N_R, the average number of level crossings per second at specified R. The level crossing rate is a function of the mobile speed as is apparent from the presence of f_m in equation (4.80). There are few crossings at both high and low levels, with the maximum rate occurring at $\rho = 1/\sqrt{2}$, (i.e., at a level 3 dB below the rms level). The signal envelope experiences very deep fades only occasionally, but shallow fades are frequent.

1

SATELLITE COMMUNICATION – AN INTRODUCTION

Contents

1.1 INTRODUCTION

- Satellites are specifically made for telecommunication purpose. They are used for mobile applications such as communication to ships, vehicles, planes, hand-held terminals and for TV and radio broadcasting.

- They are responsible for providing these services to an assigned region (area) on the earth. The power and bandwidth of these satellites depend upon the preferred size of the footprint, complexity of the traffic control protocol schemes and the cost of ground stations.

- A satellite works most efficiently when the transmissions are focused with a desired area. When the area is focused, then the emissions don't go outside that designated area and thus minimizing the interference to the other systems. This leads more efficient spectrum usage.

- Satellite's antenna patterns play an important role and must be designed to best cover the designated geographical area (which is generally irregular in shape). Satellites should be designed by keeping in mind its usability for short and long term effects throughout its life time.

- The earth station should be in a position to control the satellite if it drifts from its orbit it is subjected to any kind of drag from the external forces.

1.2 BASICS

- Satellites orbit around the earth. Depending on the application, these orbits can be circular or elliptical. Satellites in circular orbits always keep the same distance to the earth's surface following a simple law:
  The attractive force Fg of the earth due to gravity equals $m \cdot g\,(R/r)^2$
  The centrifugal force Fc trying to pull the satellite away equals $m \cdot r \cdot \omega^2$
  The variables have the following meaning:
  - m is the mass of the satellite;
  - R is the radius of earth with R = 6,370 km;
  - ri s the distance of the satellite to the centre of the earth;
  - g is the acceleration of gravity with g = 9.81 m/s2;
  - ω is the angular velocity with $\omega = 2 \cdot \pi \cdot f$, f is the frequency of the rotation.

To keep the satellite in a stable circular orbit, the following equation must hold:

$Fg = Fc$, i.e., both forces must be equal. Looking at this equation the first thing to notice is that the mass m of a satellite is irrelevant (it appears on both sides of the equation).

Solving the equation for the distance r of the satellite to the centre of the earth results in the following equation:

The distance $r = (g \cdot R^2/(2 \cdot \pi \cdot f)^2)1/3$

- From the above equation it can be concluded that the distance of a satellite to the earth's surface depends on its rotation frequency.

- Important parameters in satellite communication are the *inclination* and *elevation* angles. The inclination angle δ (figure 1.1) is defined between the equatorial plane and the plane described by the satellite orbit. An inclination angle of 0 degrees means that the satellite is exactly above the equator. If the satellite does not have a circular orbit, the closest point to the earth is called the perigee.

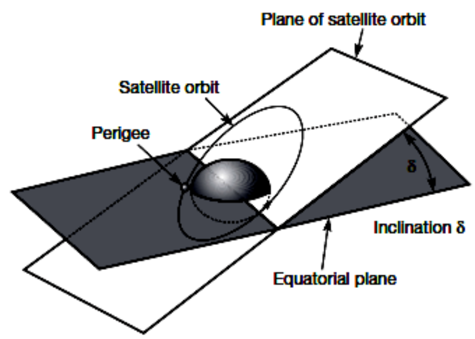


Figure 1.1: Angle of Inclination

- The elevation angle ε (figure 1.2) is defined between the centre of the satellite beam and the plane tangential to the earth's surface. A so called footprint can be defined as the area on earth where the signals of the satellite can be received.

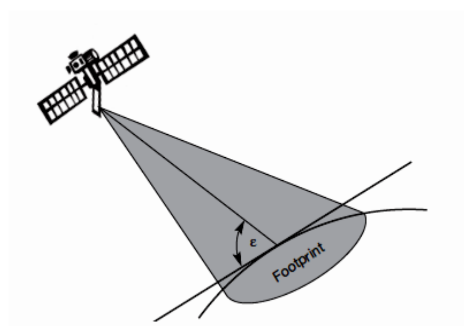


Figure 1.2: Angle of Elevation

1.3 APPLICATIONS OF SATELLITES

1.3.1) Weather Forecasting

Certain satellites are specifically designed to monitor the climatic conditions of earth. They continuously monitor the assigned areas of earth and predict the weather conditions of that region. This is done by taking images of earth from the satellite. These images are transferred using assigned radio frequency to the earth station. (Earth Station: it's a radio station located on the earth and used for relaying signals from satellites.) These satellites are exceptionally useful in predicting disasters like hurricanes, and monitor the changes in the Earth's vegetation, sea state, ocean color, and ice fields.

1.3.2) Radio and TV Broadcast

These dedicated satellites are responsible for making 100s of channels across the globe available for everyone. They are also responsible for broadcasting live matches, news, world-wide radio services. These satellites require a 30-40 cm sized dish to make these channels available globally.

1.3.3) Military Satellites

These satellites are often used for gathering intelligence, as a communications satellite used for military purposes, or as a military weapon. A satellite by itself is neither military nor civil. It is the kind of payload it carries that enables one to arrive at a decision regarding its military or civilian character.

1.3.4) Navigation Satellites

The system allows for precise localization world-wide, and with some additional techniques, the precision is in the range of some meters. Ships and aircraft rely on GPS as an addition to traditional navigation systems. Many vehicles come with installed GPS receivers. This system is also used, e.g., for fleet management of trucks or for vehicle localization in case of theft.

1.3.5) Global Telephone

One of the first applications of satellites for communication was the establishment of international telephone backbones. Instead of using cables it was sometimes faster to launch a new satellite. But, fiber optic cables are still replacing satellite communication across long distance as in fiber optic cable, light is used instead of radio frequency, hence making the communication much faster (and of course, reducing the delay caused due to the amount of distance a signal needs to travel before reaching the destination.).

Using satellites, to typically reach a distance approximately 10,000 kms away, the signal needs to travel almost 72,000 kms, that is, sending data from ground to satellite and (mostly) from satellite to another location on earth. This cause’s substantial amount of delay and this delay becomes more prominent for users during voice calls.

1.3.6) Connecting Remote Areas

Due to their geographical location many places all over the world do not have direct wired connection to the telephone network or the internet (e.g., researchers on Antarctica) or because of the current state of the infrastructure of a country. Here the satellite provides a complete coverage and (generally) there is one satellite always present across a horizon.

1.3.7) Global Mobile Communication

The basic purpose of satellites for mobile communication is to extend the area of coverage. Cellular phone systems, such as AMPS and GSM (and their successors) do not cover all parts of a country. Areas that are not covered usually have low population where it is too expensive to install a base station. With the integration of satellite communication, however, the mobile phone can switch to satellites offering world-wide connectivity to a customer. Satellites cover a certain area on the earth. This area is termed as a 'footprint' of that satellite. Within the footprint, communication with that satellite is possible for mobile users. These users communicate using a Mobile-User-Link (MUL). The base-stations communicate with satellites using a Gateway-Link (GWL). Sometimes it becomes necessary for satellite to create a communication link between users belonging to two different footprints. Here the satellites send signals to each other and this is done using Inter-Satellite-Link (ISL).

1.4 FREQUENCY ALLOCATION FOR SATELLITE

- Allocation of frequencies to satellite services s a complicated process which requires international coordination and planning. This is done as per the International Telecommunication Union (ITU). To implement this frequency planning, the world is divided into three regions:
  - Region1: Europe, Africa and Mongolia
  - Region 2: North and South America and Greenland
  - Region 3: Asia (excluding region 1 areas), Australia and south-west Pacific.
- Within these regions, he frequency bands are allocated to various satellite services. Some of them are listed below.
  - **Fixed satellite service**: Provides Links for existing Telephone Networks Used for transmitting television signals to cable companies
  - **Broadcasting satellite service**: Provides Direct Broadcast to homes. E.g. Live Cricket matches etc
  - **Mobile satellite service**s: This includes services for:
    Land Mobile
    Maritime Mobile
    Aeronautical mobile
  - **Navigational satellite services** : Include Global Positioning systems

- **Meteorological satellite services**: They are often used to perform Search and Rescue service

- Below are the frequencies allocated to these satellites:
  Frequency Band (GHZ) Designations:
  - VHF: 01-0.3
  - UHF: 0.3-1.0
  - L-band: 1.0-2.0
  - S-band: 2.0-4.0
  - C-band: 4.0-8.0
  - X-band: 8.0-12.0
  - Ku-band: 12.0-18.0 *(Ku is Under K Band)*
  - Ka-band: 18.0-27.0 *(Ka is Above K Band)*
  - V-band: 40.0-75.0
  - W-band: 75-110
  - Mm-band: 110-300
  - **μ**m-band: 300-3000
- Based on the satellite service, following are the frequencies allocated to the satellites:
  Frequency Band (GHZ) Designations:
  - VHF: 01-0.3 ---Mobile & Navigational Satellite Services
  - L-band: 1.0-2.0 --- Mobile & Navigational Satellite Services
  - C-band: 4.0-8.0 --- Fixed Satellite Service
  - Ku-band: 12.0-18.0 --- Direct Broadcast Satellite Services

1.5 TYPES OF SATELLITES (BASED ON ORBITS)

1.5.1) Geostationary or geosynchronous earth orbit (GEO)

- GEO satellites are synchronous with respect to earth. Looking from a fixed point from Earth, these satellites appear to be stationary. These satellites are placed in the space in such a way that only three satellites are sufficient to provide connection throughout the surface of the Earth (that is; their footprint is covering almost $1/3^{rd}$ of the Earth). The orbit of these satellites is circular.

- There are three conditions which lead to geostationary satellites. Lifetime expectancy of these satellites is 15 years.
  1) The satellite should be placed 37,786 kms (approximated to 36,000 kms) above the surface of the earth.
  2) These satellites must travel in the rotational speed of earth, and in the direction of motion of earth, that is eastward.
  3) The inclination of satellite with respect to earth must be 0^0.
- Geostationary satellite in practical is termed as geosynchronous as there are multiple factors which make these satellites shift from the ideal geostationary condition.

1) Gravitational pull of sun and moon makes these satellites deviate from their orbit. Over the period of time, they go through a drag. (Earth's gravitational force has no effect on these satellites due to their distance from the surface of the Earth.)
2) These satellites experience the centrifugal force due to the rotation of Earth, making them deviate from their orbit.
3) The non-circular shape of the earth leads to continuous adjustment of speed of satellite from the earth station.

- These satellites are used for TV and radio broadcast, weather forecast and also, these satellites are operating as backbones for the telephone networks.

- Disadvantages of GEO: Northern or southern regions of the Earth (poles) have more problems receiving these satellites due to the low elevation above a latitude of 60°, i.e., larger antennas are needed in this case. Shading of the signals is seen in cities due to high buildings and the low elevation further away from the equator limit transmission quality. The transmit power needed is relatively high which causes problems for battery powered devices. These satellites cannot be used for small mobile phones. The biggest problem for voice and also data communication is the high latency as without having any handovers, the signal has to at least travel 72,000 kms. Due to the large footprint, either frequencies cannot be reused or the GEO satellite needs special antennas focusing on a smaller footprint. Transferring a GEO into orbit is very expensive.

1.5.2) Low Earth Orbit (LEO) satellites:

- These satellites are placed 500-1500 kms above the surface of the earth. As LEOs circulate on a lower orbit, hence they exhibit a much shorter period that is 95 to 120 minutes. LEO systems try to ensure a high elevation for every spot on earth to provide a high quality communication link. Each LEO satellite will only be visible from the earth for around ten minutes.

- Using advanced compression schemes, transmission rates of about 2,400 bit/s can be enough for voice communication. LEOs even provide this bandwidth for mobile terminals with Omni-directional antennas using low transmit power in the range of 1W. The delay for packets delivered via a LEO is relatively low (approx 10 ms). The delay is comparable to long-distance wired connections (about 5–10 ms). Smaller footprints of LEOs allow for better frequency reuse, similar to the concepts used for cellular networks. LEOs can provide a much higher elevation in Polar Regions and so better global coverage.

- These satellites are mainly used in remote sensing an providing mobile communication services (due to lower latency).

- Disadvantages: The biggest problem of the LEO concept is the need for many satellites if global coverage is to be reached. Several concepts involve 50–200 or even more satellites in orbit. The short time of visibility with a high elevation requires additional mechanisms for connection handover between different satellites. The high number of satellites combined with the fast movements resulting in a high complexity of the whole satellite system. One general problem of LEOs is the short lifetime of about five to eight years due to atmospheric drag and radiation from the inner Van Allen belt1. Assuming 48 satellites and a lifetime of eight years, a new satellite would be needed every two months. The low latency via a single LEO is only half of the story. Other factors are the need for routing of data packets from satellite to if a user wants to communicate around the world. Due to the large footprint, a GEO typically does not need this type of routing, as senders and receivers are most likely in the same footprint.

1.5.3) Medium Earth Orbit (MEO) satellites:

- MEOs can be positioned somewhere between LEOs and GEOs, both in terms of their orbit and due to their advantages and disadvantages. Using orbits around 10,000 km, the system only requires a dozen satellites which is more than a GEO system, but much less than a LEO system. These satellites move more slowly relative to the earth’s rotation allowing a simpler system design (satellite periods are about six hours). Depending on the inclination, a MEO can cover larger populations, so requiring fewer handovers.

- Disadvantages: Again, due to the larger distance to the earth, delay increases to about 70–80 ms. the satellites need higher transmit power and special antennas for smaller footprints.

The above three are the major three categories of satellites, apart from these, the satellites are also classified based on the following types of orbits:

1.5.4) Sun- Synchronous Orbits satellites:

- These satellites rise and set with the sun. Their orbit is defined in such a way that they are always facing the sun and hence they never go through an eclipse.
- For these satellites, the surface illumination angle will be nearly the same every time.
  (Surface illumination angle: The illumination angle is the angle between the inward surface normal and the direction

of light. This means that the illumination angle of a certain point of the Earth's surface is zero if the Sun is precisely overhead and that it is 90 degrees at sunset and at sunrise.)

- Special cases of the sun-synchronous orbit are the noon/midnight orbit, where the local mean solar time of passage for equatorial longitudes is around noon or midnight, and the dawn/dusk orbit, where the local mean solar time of passage for equatorial longitudes is around sunrise or sunset, so that the satellite rides the terminator between day and night.

1.5.5) Hohmann Transfer Orbit:

This is an intermediate orbit having a highly elliptical shape. It is used by GEO satellites to reach their final destination orbits. This orbit is connected to the LEO orbit at the point of perigee forming a tangent and is connected to the GEO orbit at the point of apogee again forming a tangent.

1.5.6) Prograde orbit:

This orbit is with an inclination of less than 90°. Its direction is the same as the direction as the rotation of the primary (planet).

1.5.7) Retrograde orbit:

This orbit is with an inclination of more than 90°. Its direction is counter to the direction of rotation of the planet. Only few satellites are launched into retrograde orbit because the quantity of fuel required to launch them is much greater than for a prograde orbit. This is because when the rocket starts out on the ground, it already has an eastward component of velocity equal to the rotational velocity of the planet at its launch latitude.

1.5.8) Polar Orbits

This orbit passes above or nearly above both poles (north and south pole) of the planet on each of its revolutions. Therefore it has an inclination of (or very close to) 90 degrees. These orbits are highly inclined in shape.

1.6 EXAMPLES

1.6.1) INTELSAT

- International Telecommunication Satellite:
- Created in 1964
- Over 140 member countries
- More than 40 investing entities
- Early Bird satellite in 1965
- Six (6) evolutions of INTELSAT satellites between 1965-87
- Geostationary orbit
- Covers 3 regions:

- Atlantic Ocean Region (AOR),
- Indian Ocean Region (IOR), and
- Pacific Ocean Region (POR)

1.6.2) U.S DOMSATS

Domestic Satellite:

- In geostationary orbit
- Over 140 member countries
- Direct-to-home TV service
- Three (3) categories of U. S. DBS system: high power, medium, and low power.
- Measure in equivalent isotropic radiated power (EIRP).
- The upper limit of EIRP:
  - High power (60 dBW),
  - Medium (48 dBW), and
  - Low power (37 dBW).

1.6.3) Polar Orbiting Satellites

These satellites follow the Polar Orbits. An infinite number of polar orbits cover north and south polar regions.

- Weather (ultraviolet sensor also measure ozone level) satellites between 800 and 900 km
- National Oceanic and Atmospheric Administration (NOAA) operate a weather satellite system
- Satellite period is 102 minutes and earth rotated 25 degree.
- Estimate the sub-satellite point at the following times after the equator 90 degree E North-South crossing:
  - a) 10 minutes, 87.5 degree E and 36 degree S;
  - b) 102 minutes, 65 degree E and equator;
  - c) 120 minutes, 60 degree E and 72 degree S.
- The system uses both geostationary operational environment satellite (GOES) and polar operational environment satellite (POES)
  - Sun synchronous: they across the equator at the same local time each day
  - The morning orbit, at an altitude of 830 km, crosses the equator from south to north at 7:30 AM, and the afternoon orbit, at an altitude of 870 km, at 1:40 PM.
- Search and rescue (SAR) satellite: Cospas-Sarsat.

1.7 SUMMARY

This unit discusses the basics of satellite and elaborating the parameters which are needed to calculate the distance of an orbit to which a satellite is to be launched and the other factors which are necessary to define an orbit. Further this unit discusses the applications of satellites and elaborates on the global communication which has now become possible due to the presence of satellites. This unit also elaborates on the frequency bands used by each communication satellite.

Going further, this unit also elaborates on the types of orbits a satellite can follow to provide communication. The last segment of this unit discuses three examples of satellite revolving around the Earth.

1.8 EXERCISE

1) List the various applications of satellite.
2) Why is there a need for satellite communication?
3) List and discuss the various orbits defined for satellite communication.
4) Write a note on various satellite services available.
5) What are the primary factors needed for defining an orbit of a satellite?
6) Define elevation and inclination angles of a satellite orbit.

References

1. www.wikipedia.com
2. http://www.tech-faq.com/vsat.html
3. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
4. Rappaort, Wireless Communications Principals and Practices
5. YI Bing Lin , Wireless and Mobile Network Architectures, John Wiley
6. P. Nicopolitidis ,Wireless Networks, John Wiley
7. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication

❖❖❖❖

2

ORBITS AND LUNCHING METHODS

Contents

2.1 INTRODUCTION

- The mathematical basis o satellite orbit determination has been known since the work of Newton and Kepler in the 17th Century. Since past half century some basic laws have been applied to the man made satellites commonly known artificial satellites in the Earth's orbit.

2.2 KEPLER'S LAWS

Johann Kepler developed empirically three laws of planetary motion, based on conclusions drawn from the extensive observations of Mars by Tycho Brahe (taken around the year

1600). While they were originally defined in terms of the motion of the planets about the Sun, they apply equally to the motion of natural or artificial satellites about the Earth. Kepler'sfirst law states that the satellite follows an elliptical path in its orbit around the Earth. The satellite does not necessarily have uniform velocity around its orbit. Kepler's second law states that the line joining the satellite with the centre of the Earth sweeps out equal areas in equal times. Kepler's third law states that the cube of the mean distance of the satellite from the Earth is proportional to the square of its period.

2.2.1) Kepler's First Law

- The path followed be a satellite (in our case artificial satellite) around the primary (a planet and in our case Earth) will be an ellipse.

- "The orbit of every planer is an ellipse with sun at one of the two foci. "

- An ellipse has two focal points. Let us consider F1 and F2. The centre of mass of the two body system, known as the barycentre as always cantered at one foci. Due to the great difference between the masses of the planet (Earth) and the satellite, centre of mass always coincides with the centre of Earth and hence is always at one foci.

(Note: Ellipse: A regular oval shape, traced by a point moving in a plane so that the sum of its distances from two other points (the foci) is constant.

Foci: The center of interest and in our case centre of the ellipse.)

- Parameters associated with the 1^{st} law of Kepler:
  - ***Eccentricity (e):*** it defines how stretched out an ellipse is from a perfect circle.

  - ***Semi-Major axis (a):*** It is the longest diameter, a line that runs through the centre and both foci, its ends being at the widest points of the shapes. This line joins the points of apogee.

  - ***Semi-Minor axis (b):*** the line joining the points of perigee is called the Semi-Minor axis.

    The value of e could be determined by: $e = (\sqrt{a^2 - b^2}) / a$

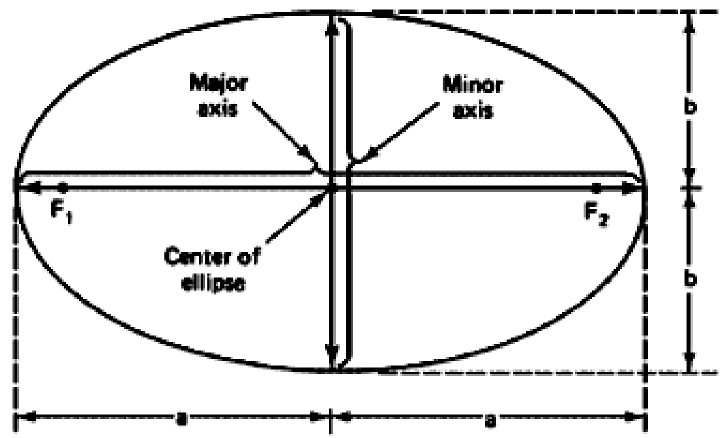


Figure 2.1: Foci F1 and F2, Semi-major axis a and semi-minor axis b of an ellipse.

<u>22.2) Kepler's Second Law</u>

- "For equal time intervals, a satellite will sweep out equal areas in its orbital plane focussed at the barycentre".

- With respect to the laws governing the planetary motion around the sun, tis law could be stated as "A line joining a planet and the sun sweeps our equal area during equal intervals of time".

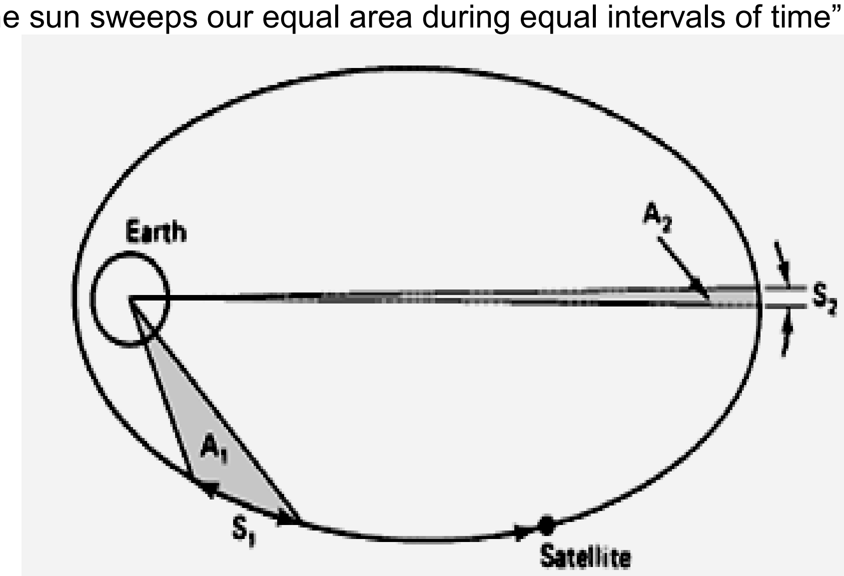


Figure 2.2: The areas A1 and A2 swept out in unit intervals of time.

- From figure 2.2 and considering the law stated above, if satellite travels distances S1 and S2 meters in 1 second, then areas A1 and A2 will be equal.

- The same area will be covered everyday regardless of where in its orbit a satellite is. As the First Keplerian law states that the satellite follows an elliptical orbit around the primary, then the satellite is at different distances from the planet at different parts of the orbit. Hence the satellite has to move faster when it is closer to the Earth so that it sweeps an equal area on the Earth.

- This could be achieved if the speed of the satellite is adjusted when it is closer to the surface of the Earth in order to make it sweep out equal areas (footprints) of the surface of the Earth.

<u>2.2.3) Kepler's Third Law</u>

- The square of the periodic time of orbit is proportional to the cube of the mean distance between the two bodies.

- The square of the orbital period of a planet is directly proportional to the cube of the semi-major axis of its orbit.

- This law shows the relationship between the distances of satellite from earth and their orbital period.

- Example: suppose satellite Satellite-I is four times as far from Earth as Satellite-II. Then **I** must traverse four times the distance of **II** in each orbit. Now considering the speed of **I** and **II**, suppose **I** travels at half the speed of **II**, then in order to maintain equilibrium with the reduced gravitational force (as **I** is four times away from Earth than what **II** is), then in all it will require 4 x 2 = 8 times as long for **I** to travel an orbit in agreement with the law which comes down to ($8^2 = 4^3$).

- Symbolically: $P^2 \alpha a^3$ (P^2 is directly proportional to a^3)
  Where **P** is the orbital period; **a** is the semi-major axis

  $$a^3 = \mu/n^2$$

  Where **n** is the mean motion of satellite in radians per second and ***μ*** is the Earth's geocentric gravitational constant.

  $$\mu = 3.986005 \times 10^{14}\ m^3/sec^2$$

Due to Earth's oblateness, a new parameter called drag is taken into account.

$$P = 2\Pi / n$$

Here, P is in seconds and n is in radians/ second

This law also confirms the fact that there is a fixed relation between period and size.

2.3 DEFINITIONS

- ***Apogee***: A point for a satellite farthest from the Earth. It is denoted as $\mathbf{h_a}$.
- ***Perigee*****:** A point for a satellite closest from the Earth. It is denoted as $\mathbf{h_p}$.
- ***Line of Apsides***: Line joining perigee and apogee through centre of the Earth. It is the major axis of the orbit. One-half of this line's length is the semi-major axis equivalents to satellite's mean distance from the Earth.
- ***Ascending Node:*** The point where the orbit crosses the equatorial plane going from north to south.
- ***Descending Node:*** The point where the orbit crosses the equatorial plane going from south to north.
- ***Inclination:*** the angle between the orbital plane and the Earth's equatorial plane. Its measured at the ascending node from the equator to the orbit, going from East to North. Also, this angle is commonly denoted as **i.**
- ***Line of Nodes:*** the line joining the ascending and descending nodes through the centre of Earth.
- ***Prograde Orbit:*** an orbit in which satellite moves in the same direction as the Earth's rotation. Its inclination is always between 0^0 to 90^0. Many satellites follow this path as Earth's velocity makes it easier to lunch these satellites.
- ***Retrograde Orbit:*** an orbit in which satellite moves in the same direction counter to the Earth's rotation.
- ***Argument of Perigee:*** An angle from the point of perigee measure in the orbital plane at the Earth's centre, in the direction of the satellite motion.
- ***Right ascension of ascending node:*** The definition of an orbit in space, the position of ascending node is specified. But as the Earth spins, the longitude of ascending node changes and cannot be used for reference. Thus for practical determination of an orbit, the longitude and time of crossing the ascending node is used. For absolute measurement, a fixed reference point in space is required. It could also be defined as "*right ascension of the ascending node; right*

ascension is the angular position measured eastward along the celestial equator from the vernal equinox vector to the hour circle of the object".

- ***Mean anamoly:*** It gives the average value to the angular position of the satellite with reference to the perigee.

- ***True anamoly:*** It is the angle from point of perigee to the satellite's position, measure at the Earth's centre.

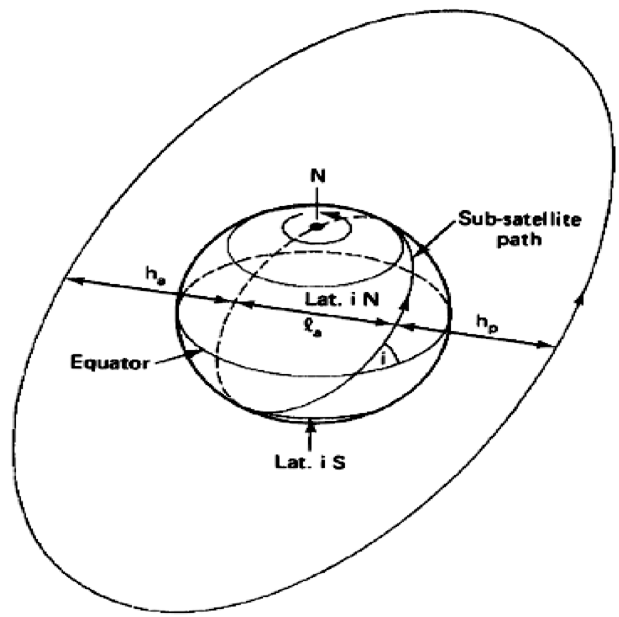


Figure 2.3: Apogee height h_a, Perigee height h_p, Inclination i, line of apsides l_a

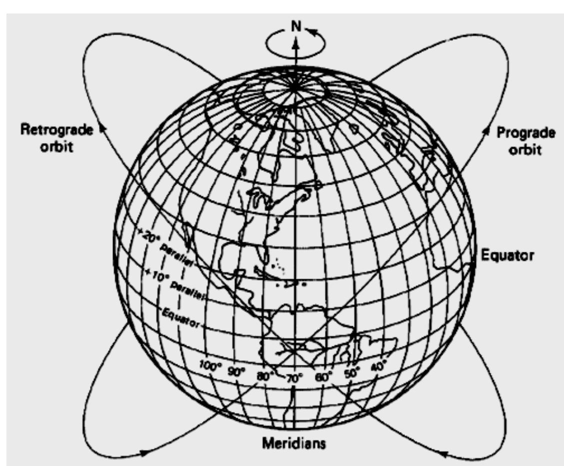


Figure 2.4: Prograde and Retrograde orbits

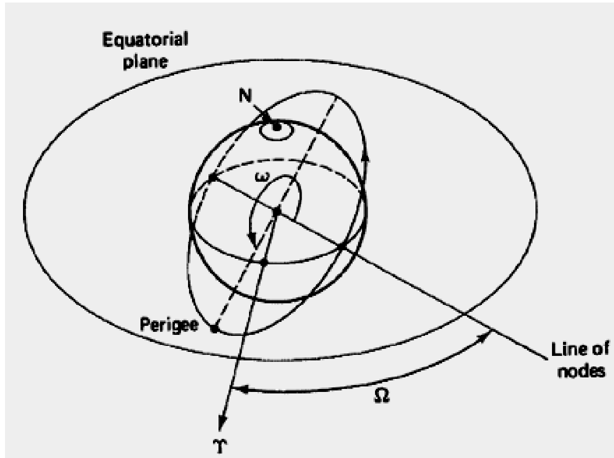


Figure 2.5: Argument of Perigee and Right ascension of ascending node

2.3.1) Orbital Elements

Following are the 6 elements of the Keplerian Element set commonly known as orbital elements.

1. Semi-Major axis (a)
2. Eccentricity (e)

   They give the shape (of ellipse) to the satellite's orbit.

3. Mean anomaly (M_0)
   It denotes the position of a satellite in its orbit at a given reference time.
4. Argument of Perigee
   It gives the rotation of the orbit's perigee point relative to the orbit's nodes in the earth's equatorial plane.
5. Inclination
6. Right ascension of ascending node

   They relate the orbital plane's position to the Earth.

As the equatorial bulge causes a slow variation in argument of perigee and right ascension of ascending node, and because other perturbing forces may alter the orbital elements slightly, the values are specified for the reference time or epoch.

2.4 ORBITAL PERBUTATIONS

- Theoretically, an orbit described by Kepler is ideal as Earth is considered to be a perfect sphere and the force acting around the Earth is the centrifugal force. This force is supposed to balance the gravitational pull of the earth.

- In reality, other forces also play an important role and affect the motion of the satellite. These forces are the gravitational forces of Sun and Moon along with the atmospheric drag.

- Effect of Sun and Moon is more pronounced on geostationary earth satellites where as the atmospheric drag effect is more pronounced for low earth orbit satellites.

2.4.1) Effects of non-Spherical Earth

- As the shape of Earth is not a perfect sphere, it causes some variations in the path followed by the satellites around the primary. As the Earth is bulging from the equatorial belt, and keeping in mind that an orbit is not a physical entity, and it is the forces resulting from an oblate Earth which act on the satellite produce a change in the orbital parameters.

- This causes the satellite to drift as a result of regression of the nodes and the latitude of the point of perigee (point closest to the Earth). This leads to rotation of the line of apsides. As the orbit itself is moving with respect to the Earth, the resultant changes are seen in the values of argument of perigee and right ascension of ascending node.

- Due to the non-spherical shape of Earth, one more effect called as the “Satellite Graveyard” is seen. The non-spherical shape leads to the small value of eccentricity (10^{-5}) at the equatorial plane. This causes a gravity gradient on GEO satellite and makes them drift to one of the two stable points which coincide with minor axis of the equatorial ellipse.

- Working satellites are made to drift back to their position but out-of-service satellites are eventually drifted to these points, and making that point a Satellite Graveyard.

  (Note: A graveyard orbit, also called a supersynchronous orbit, junk orbit or disposal orbit, is an orbit significantly above GEO where satellites are intentionally placed at the end of theiroperational life. It is a measure performed in order to lower the probability of collisions with operational spacecraft and of the generation of additional space debris. The points where the graveyard is made are separated by 1800 on the equator and are set approximately on 750 E longitude and 1050 W longitude.)

2.4.2) Atmospheric Drag

- For Low Earth orbiting satellites, the effect of atmospheric drag is more pronounces. The impact of this drag is maximumat the point of perigee. Drag (pull towards the Earth) has an effect on velocity of Satellite (velocity reduces).

- This causes the satellite to not reach the apogee height successive revolutions. This leads to a change in value of semi-major axis and eccentricity. Satellites in service are maneuvered by the earth station back to their original orbital position.

2.5 INCLINED ORBITS

- While considering an orbit of non- geostationary orbit satellite, different parameters are referred at different reference frames. The orbital elements are calculated with respect to the plane of the orbit, which is fixed in space and the earth stations position is given by geographic coordinates that rotate with the earth.

- Other factors of consideration are azimuth and elevation angles. Thus for calculation purpose, transformations between coordinate system is required.

- The following quantities and concepts are required
    - Orbital elements
    - Various measures of time
    - Perifocal coordinate system- based on orbital plane
    - Geocentric-equatorial plane coordinate system – Earth’s equatorial plane
    - Topocentric – horizon coordinate system- observer’s horizon plane.

2.5.1) Calendars

- Calendars are created with respect to the position of sun. As Earth’s motion around the sun is not uniform, its called “mean sun”. Calendar days are “mean solar days”. Tropical year has 365.2422 days. It is generally taken as 365 (commonly known as civil year).

- The extra 0.2422 is significant and for example after 100 years, there will be drift of 24 days between calendar year and tropical year. Hence the concept of Leap year came into existence.

- By the year 1582, a discrepancy was once again observed. The discrepancy existed between the civil and the tropical years. To synchronize them, days between 5th October – 14th October 1582 were abolished.

- Additional constraints were added that on years ending with two zeros to be considered as leap years. The resulting calendar is called as the Gregorian Calender; named after Pope Gregory XIII.

Example: Calculate the average length of the civil year in Gregorian calendar

Solution:
Nominal number of days in 400 years = 400 x 365
Number of leap years in 400 years = 400 / 4
This must be reduced by 3 days;
Hence, 14600 + 100 – 3 = 146097

Therefore, yearly average = 14697 / 400
= **356.2425**

2.5.2) Universal Time

Universal time coordinate (UTC) is the time used for all civic time keeping purpose. Fundamental unit of UTC is ***mean solar day***. UTC is equivalent to Greenwich mean time (GMT) and the Zullu time (Z).
1 Mean Solar Day s divided into 24 hours;
1 Hour into 60 minutes;
1 Minute into 60 seconds.
Thus there are 86,400 seconds in a day.

Example: calculate time in days, hours, minutes and seconds for epoch day 324.95616765

Solution: Mean solar day = 324^{th} + 0.95616765 mean solar day
Therefore;
24 x 0.95616765 = 22.948022
60 x 0.948022 = 56.881344
60 x 0.881344 = 52.88064

Thus; Epoch is at 22 hours 56 minutes a 52.88 seconds of the 324^{th} day of the year.

For computations, UT requires two forms:

1. Fraction of a day
2. In Degrees

2.5.3) Julian Date

- Generally time interval between two events is computed using calendar time or UT. But this notation is not suited for computations where timing of many events has to be computed.

- Thus creating a reference time in which all the events can be referred in decimal days is required. Such a time reference is provided by the Julian zero time reference, which is 12 noon (12:00 UT) on January 1, 4713 (it is a hypothetical starting point).

Example: Find the Julian Date for 13 hours UT on 18 Dec 2000

Solution:
y = 2000
mon = 12
dy = 18
hours = 13
minutes = 0
seconds = 0

d = dy + mon + hours + minutes + seconds
= 18 + 13 + 0 + 0
= 31

A = floor (y/100)
= 2000/100
= 20

B = 2 –A + floor (A/4)
= 2 – 20 + 5
= -13

C = floor (365.25 x y)
= 730500

D = floor (30.6001 x (mon + 1))
= 397.8013

JD = B + C + D + d + 1720994.5

Thus, Julian Day = 2451897.0417

Julian Time Computation

- To measure time intervals, Julian Century concept is created. A Julian century (JC) has 36525 mean solar days. The time interval is calculated with respect reference time of January 0.5, 1900 which corresponds to 2,415,020 Julian Days.

Denoting reference time as JDref, Julian century as JC and time in question as JD, then interval in JC from the reference time to the time in question is calculated as: T = (JD – JDref)/ JC

Example: Find the JD from reference time Jan 0.5 1900 to UT 13 hours of 18th December 2000

Solution:
y = 2000
mon = 12
dy = 18
hours = 13
minutes = 0
seconds = 0

d = dy + mon + hours + minutes + seconds
= 18 + 13 + 0 + 0
= 31

A = floor (y/100)
= 2000/100
= 20

B = 2 –A + floor (A/4)
= 2 – 20 + 5
= -13

C = floor (365.25 x y)
= 730500

D = floor (30.6001 x (mon + 1))
= 397.8013

JD = B + C + D + d + 1720994.5
Thus, Julian Day = 2451897.0417
JDref = 2145020
JC = 36525
Thus **T = (JD – JDref)/ JC**
= (2451897.0417 – 2415020) / 36525
T = 1.00963838 (time has no dimensions)

The Time from the referenced time to 18th Dec, 13 hours of UT is 1.00963838

2.5.4) Sidereal Time

The time measured with respect to stationary stars is called sidereal time. It is observed that one complete rotation of the Earth is relative to the fixed stars is not a complete rotation relative to the sun. This happens as Earth also moves in its orbit around the sun.

1 mean solar day = 1.0027379039 mean sidereal days
= 24 hours 3 minutes 56.5536 seconds sidereal time

2.6 SUMMARY

This unit discusses the basics of satellite and elaborating the parameters which are needed to calculate the distance of an orbit to which a satellite is to be launched and the other factors which are necessary to define an orbit. Further this unit discusses the applications of satellites and elaborates on the global communication which has now become possible due to the presence of satellites. This unit also elaborates on the frequency bands used by each communication satellite.

Going further, this unit also elaborates on the types of orbits a satellite can follow to provide communication. The last segment of this unit discuses three examples of satellite revolving around the Earth.

2.7 EXERCISE

1) List the various applications of satellite.
2) Why is there a need for satellite communication?
3) List and discuss the various orbits defined for satellite communication.
4) Write a note on various satellite services available.
5) What are the primary factors needed for defining an orbit of a satellite?
6) Define elevation and inclination angles of a satellite orbit.
7) Define the following:
   a) Apogee and Perigee
   b) Line of Apsides
   c) Ascending and Descending Nodes
   d) Line of Nodes
   e) Prograde and Retrograde Orbits
   f) Arguement of Perigee
   g) Right ascension of ascending node
   h) Mean and True anomaly

References

8. www.wikipedia.com
9. http://www.tech-faq.com/vsat.html
10. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
11. Rappaort, Wireless Communications Principals and Practices
12. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
13. P. Nicopolitidis ,Wireless Networks, John Wiley
14. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
15. http://www.britannica.com/EBchecked/topic/524891/satellite-communication

❖❖❖❖

3

GEOSTATIONARY ORBIT

Contents

3.1 INTRODUCTION

- A satellite that appears stationary with respect to Earth is hence named Geostationary. To appear stationary, these satellites have to fulfil three conditions:
    - It must travel eastward at the same rotational speed as the Earth.
    - The inclination of the orbit must be zero.
    - The orbit must be circular.

- If the satellite has to appear stationary, then it has to move at the same speed as the Earth (which is constant).

- Constant speed means equal areas must be swept out at equal intervals of time. This could only be attained using a circular orbit.

- Inclination must be zero as having any inclination would lead the satellite to move from north-south directions. Orbits with zero inclination lie in the Earth's equatorial plane.

- These satellites are 35,786 kms from the Earth. For the convenience of calculations, the value is rounded to 36,000.
- A worldwide network of operational geostationary meteorological satellites is used to provide visible and infrared images of Earth's surface and atmosphere. These satellite systems include:
    - the United States GOES
    - Meteosat, launched by the European Space Agency and operated by the European Weather Satellite Organization, EUMETSAT
    - the Japanese MTSAT
    - India's INSAT series

3.2 ANTENNA LOOK ANGLES

- The look angles for the ground station antenna are Azimuth and Elevation angles. They are required at the antenna so that it points directly at the satellite. Look angles are calculated by considering the elliptical orbit. These angles change in order to track the satellite.

- For geostationary orbit, these angels values does not change as the satellites are stationary with respect to earth. Thus large earth stations are used for commercial communications, these antennas beamwidth is very narrow and the tracking mechanism is required to compensate for the movement of the satellite about the nominal geostationary position.

- For home antennas, antenna beamwidth is quite broad and hence no tracking is essential. This leads to a fixed position for these antennas.

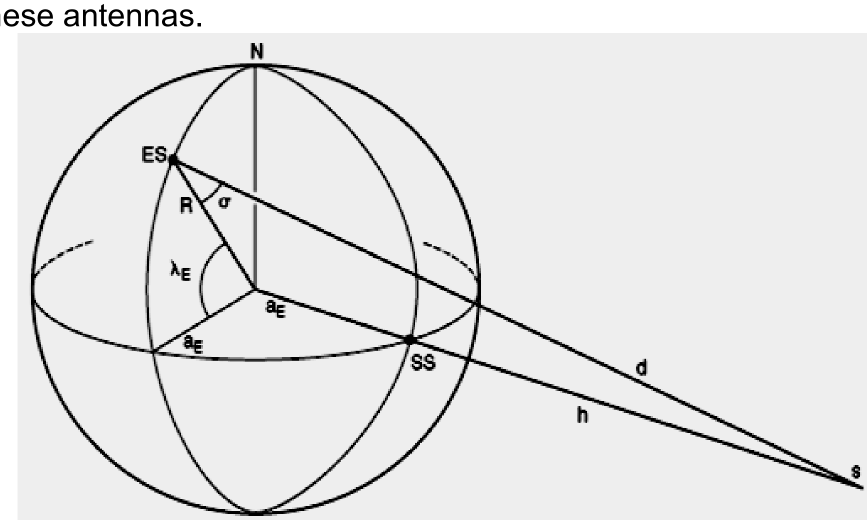


Figure 3.1: The geometry used in determining the look angles for Geostationary Satellites.

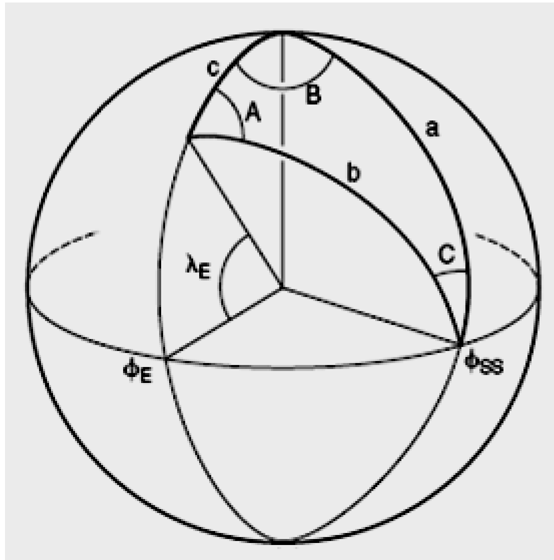


Figure 3.2: The spherical geometry related to figure 3.1

- With respect to the figure 3.1 and 3.2, the following information is needed to determine the look angles of geostationary orbit.
    1. Earth Station Latitude: λ_E
    2. Earth Station Longitude: Φ_E
    3. Sub-Satellite Point's Longitude: Φ_{SS}
    4. ES: Position of Earth Station
    5. SS: Sub-Satellite Point
    6. S: Satellite
    7. d: Range from ES to S
    8. σ: angle to be determined

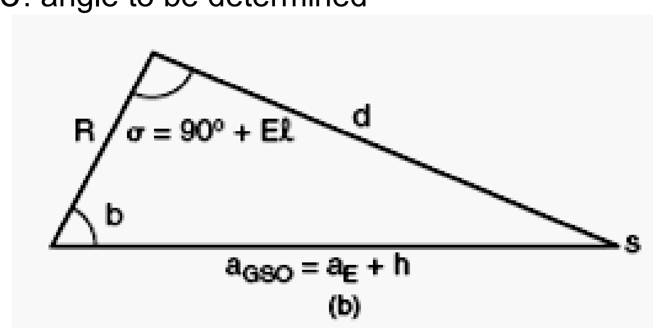


Figure 3.3: A plane triangle obtained from figure 3.1

- Considering figure 3.3, it's a spherical triangle. All sides are the arcs of a great circle. Three sides of this triangle are defined by the angles subtended by the centre of the earth.

    - Side a: angle between North Pole and radius of the sub-satellite point.

    - Side b: angle between radius of Earth and radius of the sub-satellite point.

- Side c: angle between radius of Earth and the North Pole.

  $a = 90^0$ and such a spherical triangle is called quadrantal triangle.

  $c = 90^0 - \lambda$

- Angle B is the angle between the plane containing c and the plane containing a.

  Thus, $B = \Phi_E - \Phi_{SS}$

  Angle A is the angle between the plane containing b and the plane containing c.

  Angle C is the angle between the plane containing a and the plane containing b.

  Thus, $a = 90^0$

  $c = 90^0 - \lambda_E$

  $B = \Phi_E - \Phi_{SS}$

  Thus, $b = \arccos(\cos B \cos \lambda_E)$

  $A = \arcsin(\sin |B| / \sin b)$

Example: A geostationary satellite is located 90^0 W. Calculate the azimuth angle for an Earth station antenna located at latitude 35^0 W and longitude 100^0 W.

Solution: The given quantities are:

$\Phi_E = - 100$ degrees;

$\Phi_{SS} = - 90$ degrees;

$\boldsymbol{\lambda_E} = 35$ degrees

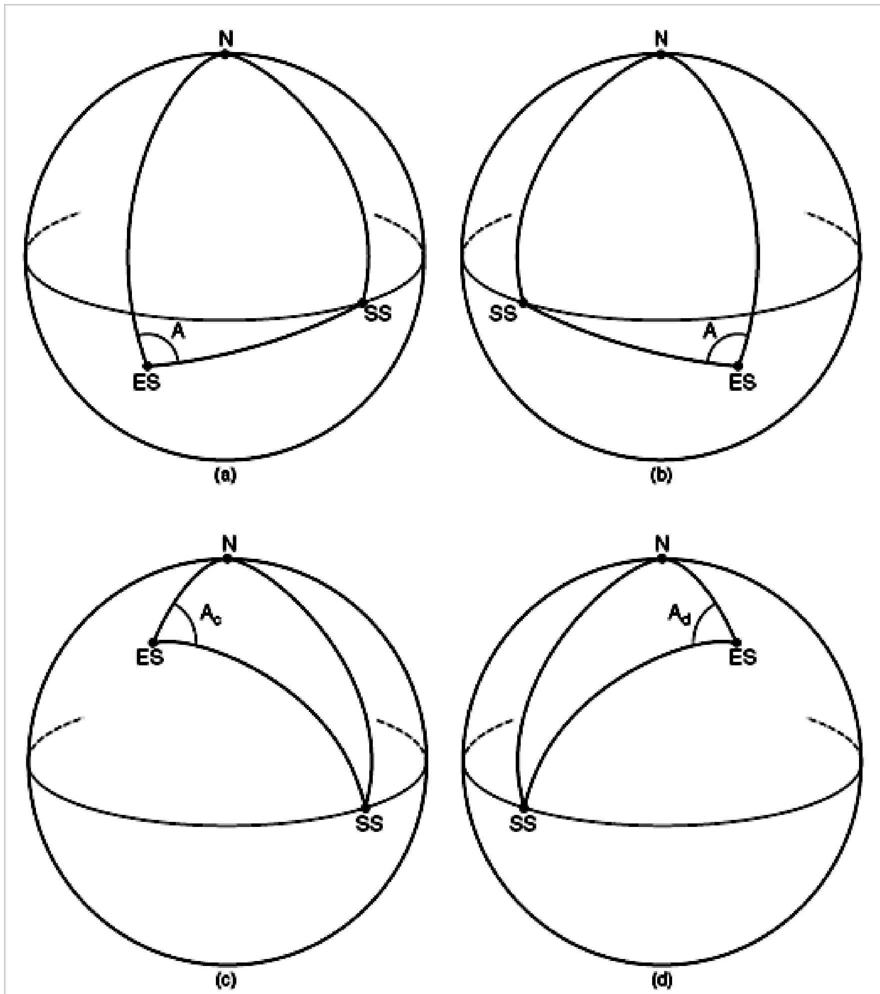


Figure 3.4: Azimuth angle related to angle A with respect to table 1.

Fig. 3.3	λ_E	B	A_z, degrees
a	<0	<0	A
b	<0	>0	$360° - A$
c	>0	<0	$180° - A$
d	>0	>0	$180° + A$

Table 1: Azimuth Angle Az from figure 3.3

$$B := \phi_E - \phi_{SS} \qquad B = -10 \cdot deg$$

$$b := acos(\cos(B) \cdot \cos(\lambda_E)) \qquad b = 36.2 \cdot deg$$

$$A := asin\left(\frac{\sin(|B|)}{\sin(b)}\right) \qquad A = 17.1 \cdot deg$$

By inspection, $\lambda_E > 0$ and B < 0. Therefore Figure 3.4 (c) applies, and;

$$A_z := 180 \cdot deg - A \qquad A_z = 162.9 \cdot deg$$

Applying the cosine rule for plane triangle to the triangle of figure 3.3 allows the range d to be found to a close approximation:

$$d = \sqrt{R^2 + a_{GSO}{}^2 - 2Ra_{GSO}\cos b}$$

Applying the sine rule for plane triangles to the triangle of figure 3.3 allows the angle of elevation to be found:

$$El = \arccos\left(\frac{a_{GSO}}{d}\sin b\right)$$

3.3 POLAR MOUNT ANTENNAS

- These antennas are pointing accurately only for one satellite. They have a single actuator which moves the antenna in a circular arc. Generally some pointing error is seen in these antennas. The dish of this antenna is mounted on an axis termed as polar axis such that the antenna bore sight is normal to this axis. As in figure 3.5.

- The angle between polar mount and the local horizontal plane is set equal to the earth station latitude λ_E, making bore sight lie parallel to the equatorial plane. Now the axis is tilted at an angle S, which is relative to the polar mount until the bore sight is pointing at a satellite position.

$$\delta = 90^0 - El_0 - \lambda_E$$

Where El_0 is the elevation required for the satellite position.

Thus $\cos El_0 = (a_{gso} / d) \sin \lambda_E$

Hence $\delta = 900 - \arccos [(a_{gso} / d) \sin \lambda_E] - \lambda_E$

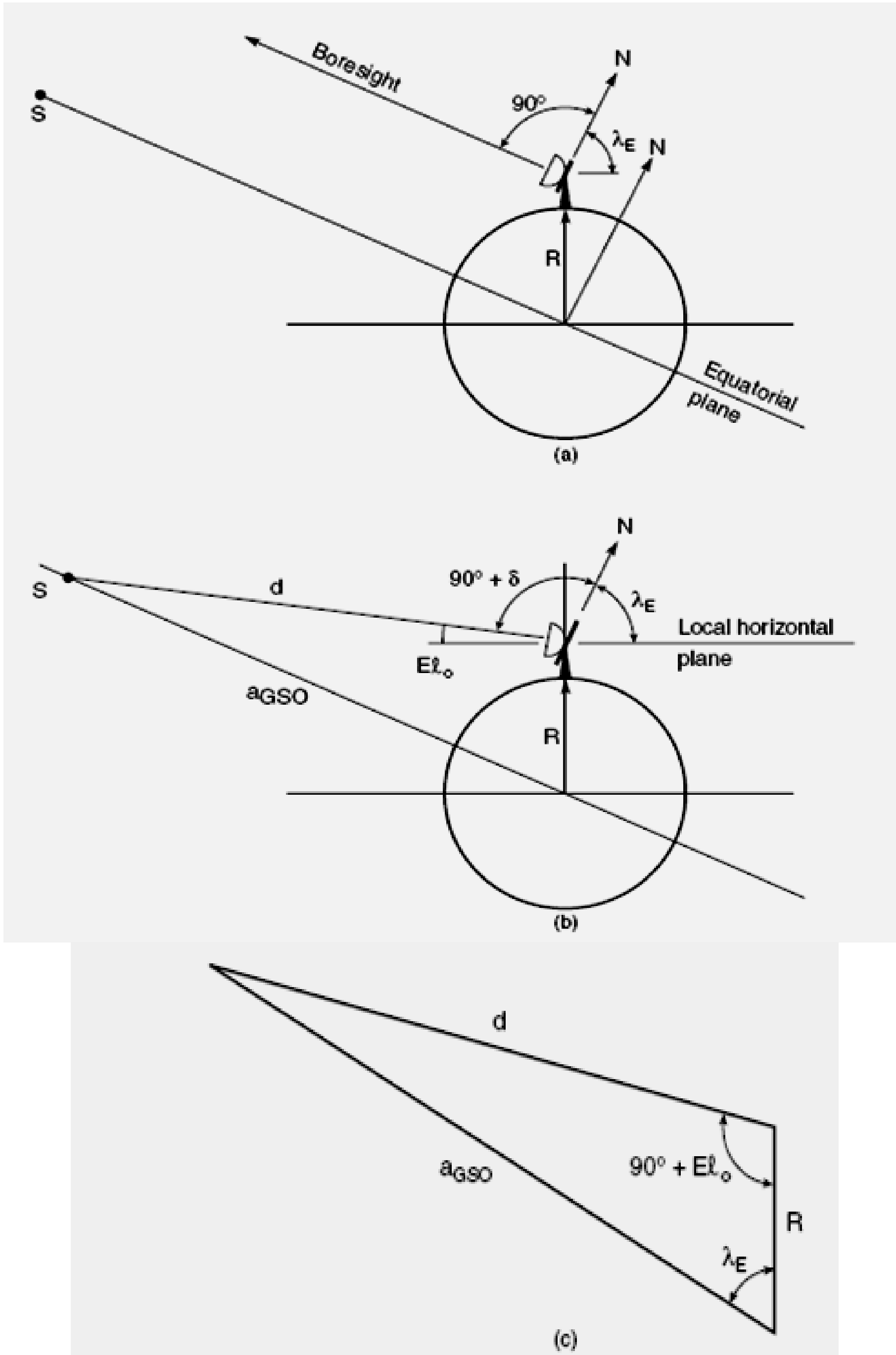


Figure 3.6 (a), (b) and (c): Polar mount Antenna

Example: Determine the angle of tilt required for polar mount used with an earth station at latitude 49 degrees north. Assume a spherical Earth of mean radius 6371 km, and ignore the earth station latitude.

***Solution*:**

$$\lambda_E := 49 \cdot deg \qquad a_{GSO} := 42164 \cdot km \qquad R := 6371 \cdot km$$

$$d := \sqrt{R^2 + a_{GSO}{}^2 - 2 \cdot R \cdot a_{GSO} \cdot \cos(\lambda_E)}$$

$$El_0 := acos\left(\frac{a_{GSO}}{d} \cdot \sin(\lambda_E)\right)$$

$$\delta := 90 \cdot deg - El_0 - \lambda_E \qquad \delta = 7 \cdot deg$$

3.4 LIMITS OF VISIBILITY

The east and west limits of geostationary are visible from any given Earth station. These limits are set by the geographic coordinates of the Earth station and antenna elevation. The lowest elevation is zero (in theory) but in practice, to avoid reception of excess noise from Earth. Some finite minimum value of elevation is issued. The earth station can see a satellite over a geostationary arc bounded by **+- (81.30)** about the earth station's longitude.

3.5 NEAR GEOSTATIONARY ORBITS

- There are a number of perbuting forces that cause an orbit to depart from ideal Keplerian orbit. The most effecting ones are gravitational fields of sun and moon, non-spherical shape of the Earth, reaction of the satellite itself to motor movements within the satellites.

- Thus the earth station keeps manoeuvring the satellite to maintain its position. Within a set of nominal geostationary coordinates. Thus the exact GEO is not attainable in practice and the orbital parameters vary with time. Hence these satellites are called "Geosynchronous" satellites or "Near-Geostationary satellites".

3.6 EARTH ECLIPSE OF A SATELLITE

- It occurs when Earth's equatorial plane coincides with the plane f he Earth's orbit around the sun. Near the time of spring and autumnal equinoxes, when the sun is crossing the equator, the satellite passes into sun's shadow. This happens for some duration of time every day.

- These eclipses begin 23 days before the equinox and end 23 days after the equinox. They last for almost 10 minutes at the beginning and end of equinox and increase for a maximum period of 72 minutes at a full eclipse. The solar cells of the satellite become non-functional during the eclipse period and the satellite is made to operate with the help of power supplied from the batteries.

- A satellite will have the eclipse duration symmetric around the time t=Satellite Longitude/15 + 12 hours. A satellite at Greenwich longitude 0 will have the eclipse duration symmetric around 0/15 UTC +12hours = 00:00 UTC. The eclipse will happen at night but for satellites in the east it will happen late evening local time. For satellites in the west eclipse will happen in the early morning hour's local time. An earth caused eclipse will normally not happen during peak viewing hours if the satellite is located near the longitude of the coverage area. Modern satellites are well equipped with batteries for operation during eclipse.

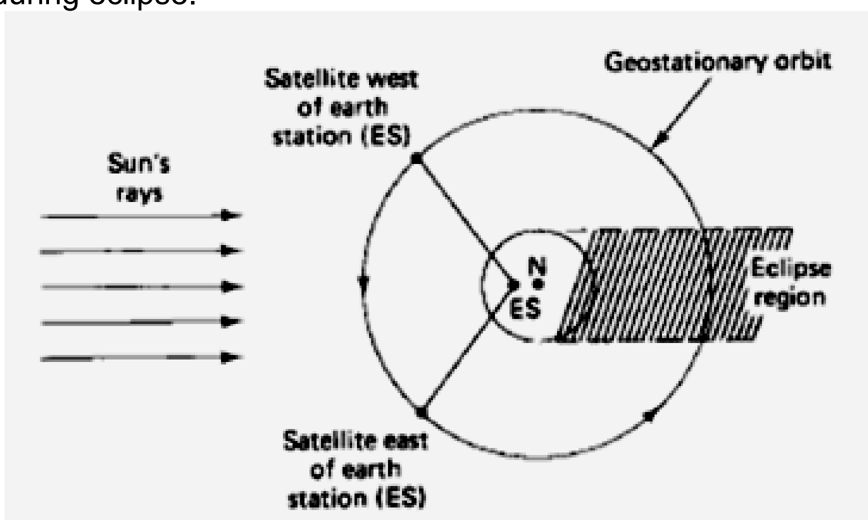


Figure 3.7: A satellite east of the earth station enters eclipse during daylight busy) hours at the earth station. A Satellite west of earth station enters eclipse during night and early morning hours (non busy time).

3.7 SUN TRANSIT OUTAGE

- Sun transit outage is an interruption in or distortion of geostationary satellite signals caused by interference from solar radiation. Sun appears to be an extremely noisy source which completely blanks out the signal from satellite. This effect lasts for 6 days around the equinoxes. They occur for a maximum period of 10 minutes.

- Generally, sun outages occur in February, March, September and October, that is, around the time of the equinoxes. At these times, the apparent path of the sun across the sky takes it directly behind the line of sight between an earth station and a satellite. As the sun radiates strongly at the microwave frequencies used to communicate with satellites (C-band, Ka band and Ku band) the sun swamps the signal from the satellite.

- The effects of a sun outage can include partial degradation, that is, an increase in the error rate, or total destruction of the signal.

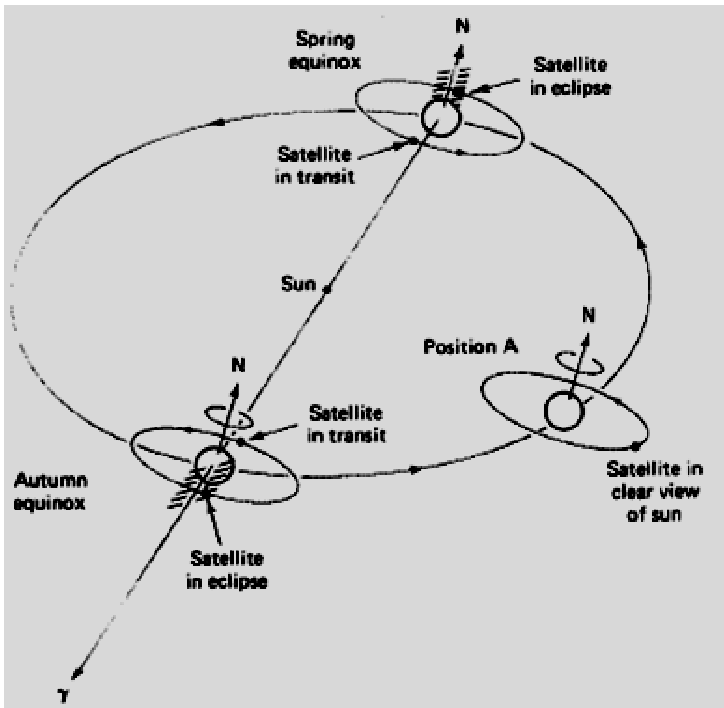


Figure 3.8: Earth Eclipse of a Satellite and Sun transit Outage

3.8 LAUNCHING ORBITS

- Low Earth Orbiting satellites are directly injected into their orbits. This cannot be done incase of GEOs as they have to be positioned 36,000kms above the Earth's surface. Launch vehicles are hence used to set these satellites in their orbits. These vehicles are reusable. They are also known as 'Space Transportation System' (STS).

- When the orbital altitude is greater than 1,200 km it becomes expensive to directly inject the satellite in its orbit. For this purpose, a satellite must be placed in to a transfer orbit between the initial lower orbit and destination orbit. The transfer orbit is commonly known as *Hohmann-Transfer Orbit.

- (*About Hohmann Transfer Orbit: This manoeuvre is named for the German civil engineer who first proposed it, Walter Hohmann, who was born in 1880. He didn't work in rocketry professionally (and wasn't associated with military rocketry), but was a key member of Germany's pioneering Society for Space Travel that included people such as Willy Ley, Hermann, and Werner von Braun. He published his concept of how to transfer between orbits in his 1925 book, The Attainability of Celestial Bodies.)

- The transfer orbit is selected to minimize the energy required for the transfer. This orbit forms a tangent to the low attitude orbit at the point of its perigee and tangent to high altitude orbit at the point of its apogee.

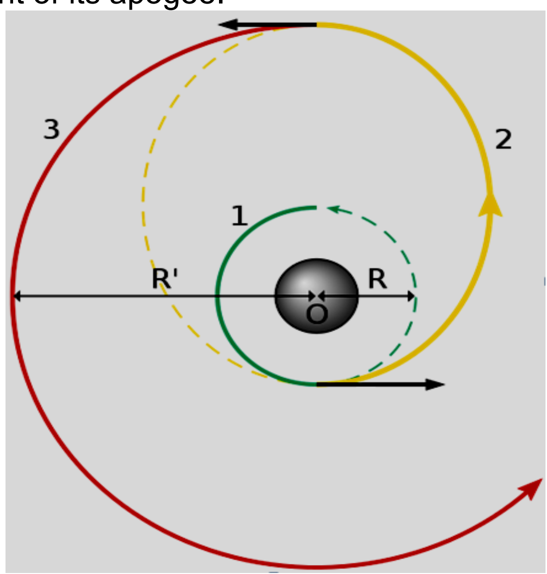


Figure 3.9: Orbit Transfer positions

- The rocket injects the satellite with the required thrust** into the transfer orbit. With the STS, the satellite carries a perigee kick motor*** which imparts the required thrust to inject the satellite in its transfer orbit. Similarly, an apogee kick motor (AKM) is used to inject the satellite in its destination orbit.

- Generally it takes 1-2 months for the satellite to become fully functional. The Earth Station performs the Telemetry Tracking and Command**** function to control the satellite transits and functionalities.

  (**Thrust: It is a reaction force described quantitatively by Newton's second and third laws. When a system expels or accelerates mass in one direction the accelerated mass will cause a force of equal magnitude but opposite direction on that system.)

(***Kick Motor refers to a rocket motor that is regularly employed on artificial satellites destined for a geostationary orbit. As the vast majority of geostationary satellite launches are carried out from spaceports at a significant distance away from Earth's equator, the carrier rocket would only be able to launch the satellite into an elliptical orbit of maximum apogee 35,784-kilometres and with a non-zero inclination approximately equal to the latitude of the launch site.)

(****TT&C: it's a sub-system where the functions performed by the satellite control network to maintain health and status, measure specific mission parameters and processing over time a sequence of these measurement to refine parameter knowledge, and transmit mission commands to the satellite. Detailed study of TT&C in the upcoming units.)

- It is better to launch rockets closer to the equator because the Earth rotates at a greater speed here than that at either pole. This extra speed at the equator means a rocket needs less thrust (and therefore less fuel) to launch into orbit. In addition, launching at the equator provides an additional 1,036 mph (1,667 km/h) of speed once the vehicle reaches orbit. This speed bonus means the vehicle needs less fuel, and that freed space can be used to carry more pay load.

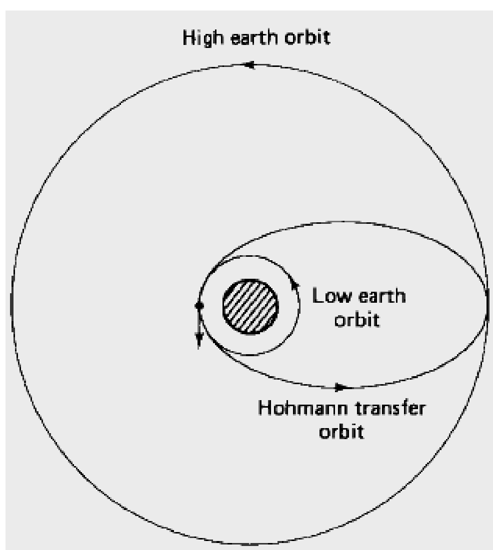


Figure 3.10: Hohmann Transfer Orbit

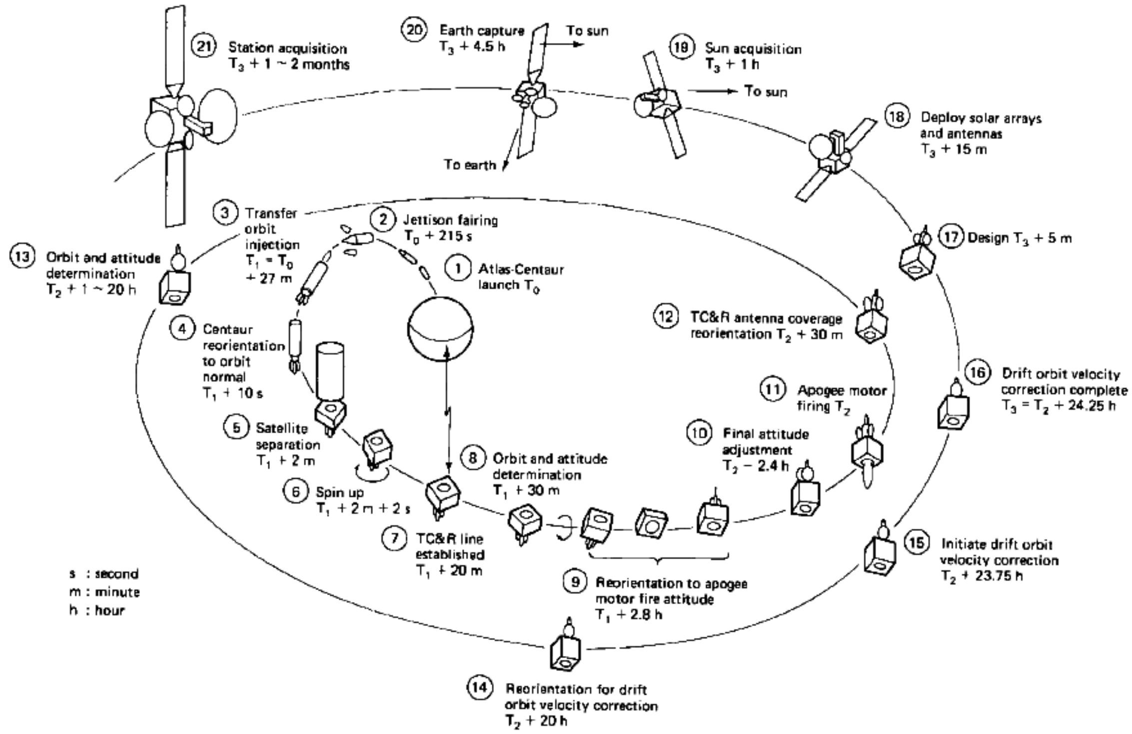


Figure 3.10: Launching stages of a GEO (example INTELSAT)

3.9 SUMMARY

- A geostationary orbit, or Geostationary Earth Orbit (GEO), is a circular orbit 35,786 km above the Earth's equator and following the direction of the Earth's rotation. An object in such an orbit has an orbital period equal to the Earth's rotational period (one sidereal day), and thus appears motionless, at a fixed position in the sky, to ground observers. Communications satellites and weather satellites are often given geostationary orbits, so that the satellite antennas that communicate with them do not have to move to track them, but can be pointed permanently at the position in the sky where they stay. A geostationary orbit is a particular type of geosynchronous orbit.

- A geostationary orbit can only be achieved at an altitude very close to 35,786 km (22,236 mi), and directly above the Equator. This equates to an orbital velocity of 3.07 km/s or a period of 1,436 minutes. This ensures that satellite is locked to the Earth's rotational period and has a stationary footprint on the ground. All geostationary satellites have to be located on this ring.

- Satellites in geostationary orbit must all occupy a single ring above the Equator. The requirement to space these satellites apart to avoid harmful radio-frequency interference during operations means that there are a limited number of orbital "slots" available, thus only a limited number of satellites can be operated in geostationary orbit. This has led to conflict between different countries wishing access to the same orbital slots (countries near the same longitude but differing latitudes) and radio frequencies.

- Further this chapter speaks about Antenna Look Angles which are the elevation and azimuth at which a particular satellite is predicted to be found at a specified time.

- And polar mount antennas which are designed to allow all visible geostationary satellites to be accessed by swinging the antenna around one axis (the main axis).

- To conclude with, how a GEO is launched using a transfer orbit is explained. Transfer orbit: Hohmann transfer orbit to bring a spacecraft from a lower circular orbit into a higher one. It is one half of an elliptic orbit that touches both the lower circular orbit that one wishes to leave and the higher circular orbit that one wishes to reach. The transfer is initiated by firing the spacecraft's engine in order to accelerate it so that it will follow the elliptical orbit; this adds energy to the spacecraft's orbit. When the spacecraft has reached its destination orbit, its orbital speed (and hence its orbital energy) must be increased again in order to change the elliptic orbit to the larger circular one.

3.10 EXERCISE

1) What is a geostationary orbit?
2) Which conditions should be fulfilled to attain a geostationary orbit?
3) Why is it more prudent to launch a satellite from a point closer to an equator?
4) Why the term ‘geosynchronous’ is used instead of ‘geostationary’?
5) How are GEO satellites made to inject to different orbits?
6) Discuss the phenomenon of Sun Transit Outage.
7) What are limits of visibility?
8) How does a satellite go through an eclipse? Explain with appropriate diagram.
9) How is a geostationary satellite launched to a GEO?
10) Does the motion of Earth beneficial for GEO satellites? Justify.
11) List the advantages and disadvantages of GEO satellites.
12) What are Look angles? Derive an equation for the same.
13) Write a note on Polar Mount Antennas.
14) It is preferable to operate with a satellite positioned at West rather than East of Earth station longitude? Justify

References

16. www.wikipedia.com
17. http://www.tech-faq.com/vsat.html
18. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
19. Rappaort, Wireless Communications Principals and Practices
20. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
21. P. Nicopolitidis ,Wireless Networks, John Wiley
22. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
23. http://www.britannica.com/EBchecked/topic/524891/satellite-communication

❖❖❖❖

4

RADIO WAVE PROPAGATION

Contents

4.1 INTRODUCTION

Radio propagation is the behavior of radio waves when they are transmitted, or propagated from one point on the Earth to another, or into various parts of the atmosphere. As a form of electromagnetic radiation, like light waves, radio waves are affected by the phenomena of reflection, refraction, diffraction, absorption, polarization and scattering.

Radio propagation is affected by the daily changes of water vapor in the troposphere and ionization in the upper atmosphere, due to the Sun. Understanding the effects of varying conditions on radio propagation has many practical applications, from choosing frequencies for international shortwave broadcasters, to designing reliable mobile telephone systems, to radio navigation, to operation of radar systems.

4.2 ATMOSPHERIC LOSSES

- Multiple losses occur due to the Earth's atmosphere. Losses maybe because of the adverse weather conditions or because of the energy absorption done by the various gases present in the atmosphere. Weather related losses are called "atmospheric attenuation". Absorption losses are called "atmospheric absorption".

- At various frequencies, different components of atmosphere cause impairments to the radio wave signals.

- Example: water vapour at 22.3 GHz and oxygen (O_2) at60 GHz.

- Considering the elevation angle of signals as 100 and absorption loss at $[AA]_{90}$ decibels. Formula for absorption loss is:

  $$[AA] = [AA]_{90} \operatorname{cosec} \theta$$

  Where θ is the elevation angle (as described in Unit III)

- A fading phenomenon, which causes the radio waves to focus and defocus because of the differences in the atmospheric refraction index is seen. This effect is called "atmospheric scintillation".

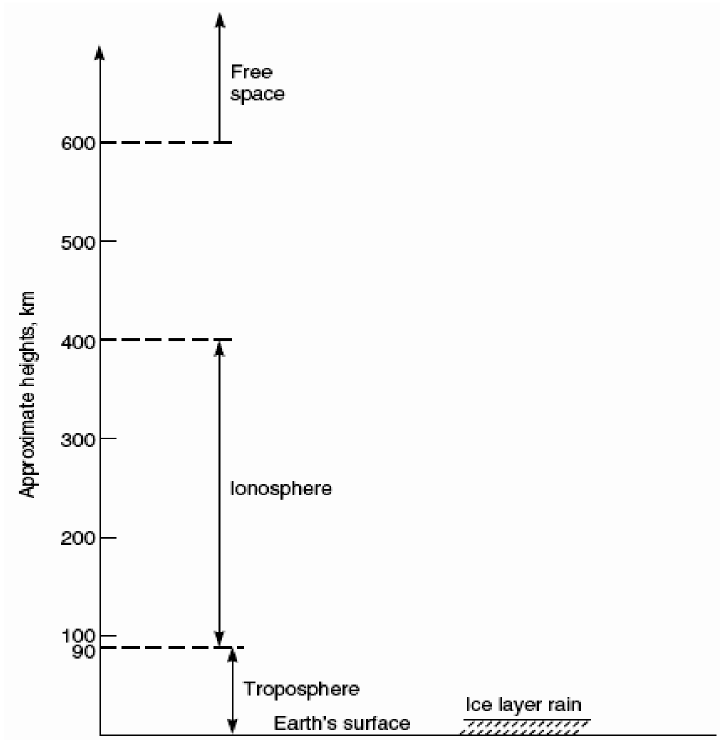


Figure 4.1: Layers of Earth's atmosphere

4.3 IONOSPHERE EFFECT

- Ionosphere is one of the layers in the Earth's atmosphere. It is situated between 90 kms to 400 kms above the surface of the Earth. All the communication signals between satellites and earth stations have to pass through this layer.

- This layer contains free electrons which are charged due to solar radiation. These ions are not uniformly distributed across the ionosphere, but move together across the ionosphere in *clusters*. Such clusters are called clouds of electrons or "travelling ionosphere disturbances". When signals pass through such electron clouds, fluctuations are caused.

- Electron clouds are created when accelerated charged particles disturb stray electrons already floating in the atmosphere, and bounce or slingshot the electrons into each other and the passing by signals. These stray electrons can be photo-electrons from synchrotron radiation or electrons from ionized gas molecules and have adverse effect on the signals passing through them especially if the density of these clouds is high.

- The other effects seen on the signal also includes scintillation, absorption, propagation delay, dispersion, and frequency change and polarization rotation. These effects decrease as the frequency increases. Out of the above effect only scintillation and polarization rotation are of major concern for satellite communication.

- ***Absorption***: Electromagnetic waves are absorbed in the atmosphere according to wavelength. Two compounds are responsible for the majority of signal absorption: oxygen (O2) and water. It is seen for frequencies at 22 GHz due to water, and at 63 GHz due to oxygen. The concrete amount of water vapour and oxygen in the atmosphere normally declines with an increase in altitude because of the decrease in pressure.

- ***Propagation Delay***: Propagation delay is the time required for a signal to travel from the sender (in our case from an earth station or a spacecraft) to the receiver. It is measured in microsecond. As distances are very much greater than those involved with terrestrial systems, propagation delay can be an issue, especially for satellites using geostationary orbits. Here the round trip from the ground to the satellite and back can be of the order of a quarter of a second.

- ***Dispersion***: Here the signals are distributed over a wide area.

- ***Polarization Rotation***: It is the phenomenon in which waves of light or other radiation are restricted in direction of vibration.

- ***Scintillation***: it is the variation in the amplitude, phase, polarization, angle of arrival of radio waves. They are caused by the irregularities in the ionosphere which change with time. Fading of signal is the major effect o ionosphere scintillation. The effect of fading can sometimes be very severe and may last upto several minutes.

Propagation impairment	Physical cause	Prime importance
Attenuation and sky noise increases	Atmospheric gases, cloud, rain	Frequencies above about 10 GHz
Signal depolarization	Rain, ice crystals	Dual-polarization systems at C and Ku bands (depends on system configuration)
Refraction, atmospheric multipath	Atmospheric gases	Communication and tracking at low elevation angles
Signal scintillations	Tropospheric and ionospheric refractivity fluctuations	Tropospheric at frequencies above 10 GHz and low elevation angles; ionospheric at frequencies below 10 GHz
Reflection multipath, blockage	Earth's surface, objects on surface	Mobile satellite services
Propagation delays, variations	Troposphere, ionosphere	Precise timing and location systems; time-division multiple access (TDMA) systems
Intersystem interference	Ducting, scatter, diffraction	Mainly C band at present; rain scatter may be significant at higher fre quencies

Table 4.1: Propagation concerns of Satellite Communication Systems

4.4 RAIN ATTENUATION

The rate at which the rain water would get accumulated in a rain gauge in the area of interest is called rain rate. Rain attenuation is a function of rain rate. It is calculated in percentage time. A percentage time is generally of a year.

Example: Rain rate of 0.001 percent means the rain rate would exceed by 0.001 percent of a year and it is be denoted as $R_{0.001}$.
Percentage time is denoted as p and rain rate as R_p.
Specific attenuation α is given by:

$$\alpha = a\, R^b_p \text{dB/km}$$

Where a and b depend upon frequency and polarization.
Total attenuation, denoted by A is given by:

$$A = \alpha\, L \quad \text{dB}$$

Where L is the effective path length of the signal through rain.

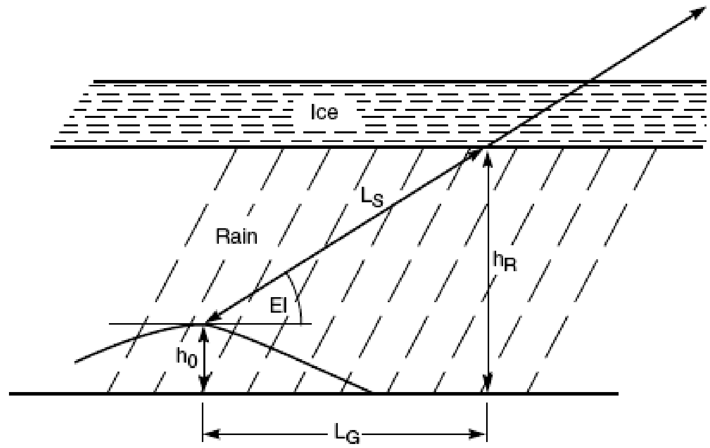


Figure 4.2: Path length through rain

Here, Ls -> Slant Height and it depends upon the antenna angle of elevation θ and rain height.
h_R -> Rain Height at which freezing occurs.
L_G -> Horizontal projection of Ls
Thus effective path length L is given by:

$$L = Ls\ r_p$$

Where $r_p$is the reduction factor of percentage time p andL_G

That is; $L_G = Ls\cos El$

Therefore rain attenuation is$\mathbf{A_p = a\ R^b{}_p\ Ls\ r_p}$dB

4.5 OTHER IMPAIRMENTS

Due to the low water content in them, rain, ice and hail have a little effect on attenuation. Attenuation can be caused by clouds but generally its effect is comparatively low.

4.6 SUMMARY

This unit discusses how the atmosphere surrounding the earth affects the signals which pass through it. The layer of Ionosphere is majorly responsible for degrading the satellite signal. The ionosphere losses that mainly occur due to the electron clouds are pre calculated by the earth stations and accordingly the frequency of signal transmitted is adjusted. Further this Unit explains the effect of rain on a signal and calculates the rain attenuation.

4.6 EXERCISE

1. Explain the phenomenon of scintillation.
2. What are atmospheric losses?
3. Explain the effects of clouds of electrons.
4. List various propagation concerns for satellite communication systems.
5. Explain ionosphere scintillation.
6. What is meant by rain attenuation? Derive an equation for the same.
7. Explain what is meant by effective path length in connection with rain attenuation.
8. Explain what is meant by rain rate. How this is related to specific attenuation.
9. Describe the major effect that ionosphere has on satellite signals losses?

References

24. www.wikipedia.com
25. http://www.tech-faq.com/vsat.html
26. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
27. Rappaort, Wireless Communications Principals and Practices
28. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
29. P. Nicopolitidis ,Wireless Networks, John Wiley
30. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
31. http://www.britannica.com/EBchecked/topic/524891/satellite-communication

❖❖❖❖

5

POLARIZATION

Contents:

5.1 INTRODUCTION

Definitions*:*

- ✓ ***Polarization***: it is a property that describes the orientation of a signal with respect to time-varying direction and amplitude of an electromagnetic wave with respect to electric field vector.

- ✓ ***Amplitude***: it is the maximum point of a vibration or oscillation, measured from the position of equilibrium.

- ✓ ***Electromagnetic wave***: The wave of the electric field and the wave of the magnetic field are propagated perpendicularly to the direction of propagation and to each other. At extremely low frequencies, the electric field and the magnetic field are specified separately. At higher frequencies, electric and magnetic fields are not separable, and are named electromagnetic waves.

- ✓ ***Vector:***Aquantity having direction as well as magnitude, especially while determining the position of one point in space relative to another.

- ✓ ***Right-Hand Thumb Rule***: For a current-carrying wire, the rule that if the fingers of the right hand are placed around the wire so that the thumb points in the direction of current flow, the fingers will be pointing in the direction of the magnetic field produced by the wire. Also known as hand rule. For a moving wire in a magnetic field, such as the wire on the armature of a generator, if the thumb, first, and second fingers of the right hand are extended at right angles to one another, with the first finger representing the direction of magnetic lines of force and the second finger representing the direction of current flow induced by the wire's motion, the thumb will be pointing in the direction of motion of the wire.

- ✓ ***Dipole***:A pair of equal and oppositely charged or magnetized poles separated by a distance.

- ✓ ***Co-Polar Component:***
- ✓ ***Torque:***it's a tendency of force to rotate an object about its axis or fulcrum or pivot. It's the cross product between distance vector and force vector.

5.2 PLANE TRANSVERSE ELECTROMAGNETIC WAVE (TEM)

- The signal sent to a satellite is generally in the form of a Transverse Electromagnetic Wave. In TEM wave electric field vector is perpendicular to the direction of propagation. In the figure, magnetic field H and electric field E are transverse to the direction of propagation that is vector k.

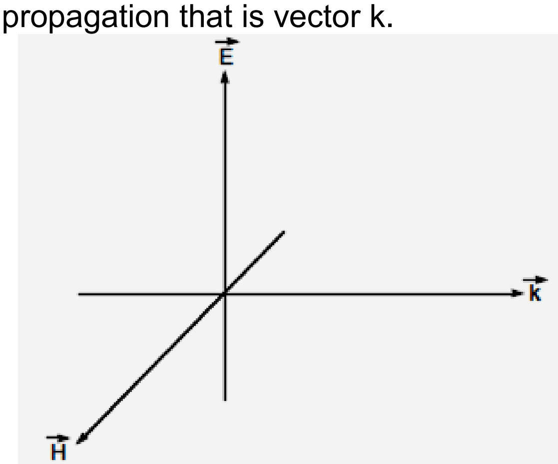


Figure 5.1: Vector diagram for TEM wave

- E, H and k are vector quantities. The path defined when rotation from E to H in the direction of rotation of right hand threaded screw, they form a right hand set. Considering TEM wave as a plane, then vectors E and H lie in the plane which is at right angles to vector k. Hence k is normal to the plane.

- The direction of the line traced out by the tip of the electric field vector determines the polarization of the wave.

- Electric and magnetic fields are varying functions of time. Magnetic field varies exactly in phase with the electric field and its amplitude is proportional to electric field's amplitude. Thus only electric field is considered.

- Tim of vector E traces out a straight line also referred as linear polarization. When E is perpendicular to the Earth's surface, it's called vertical polarization and when it's perpendicular to Earth's surface, it's called horizontal polarization.

- Let horizontal moment be x and vertical be y axes of right hand set.
  Thus vertical polarized electric field can be described as

$$Ey = \hat{a}y\ Ey \sin \omega t$$

Where: ây is unit vector in vertical direction, Ey is the peak value (magnitude) of electric field.

Similarly Horizontal polarization id

$$\mathbf{Ex = \hat{a}x\ Ex \sin \omega t}$$

- When both these fields are present together, the resultant vector is at

$$\boldsymbol{\alpha = \arctan Ey / Ex}$$

Ey and Ex are orthogonal (at right angles) but lead each other by 90^0in phase.

$$\mathbf{Ey = \hat{a}y\ E \sin \omega t}$$
$$\mathbf{Ex = \hat{a}x\ E \cos \omega t}$$

Hence, $\boldsymbol{\alpha = \omega t}$

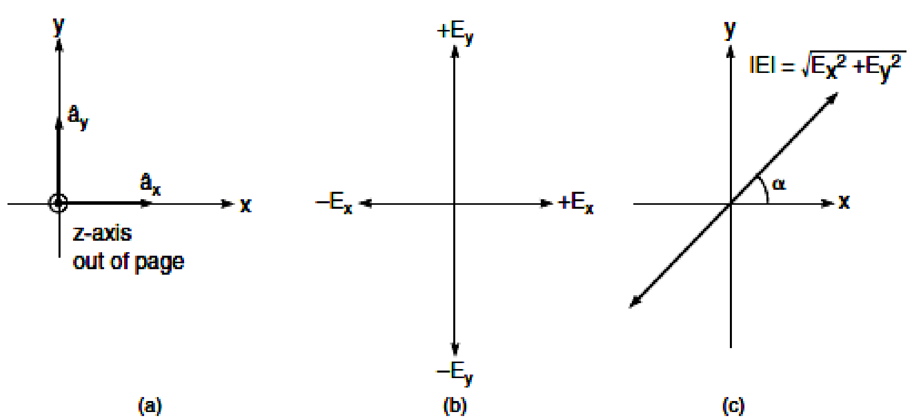


Figure 5.2: Horizontal and Vertical components of a linear polarization

- When tip of the resultant vector traces out a circle, the resultant wave is circularly polarized whose direction is defined by the rotation of the vector. It can follow right-hand circular (RHC) clock-wise direction or left-hand circular (LHC) anti-clock wise direction.

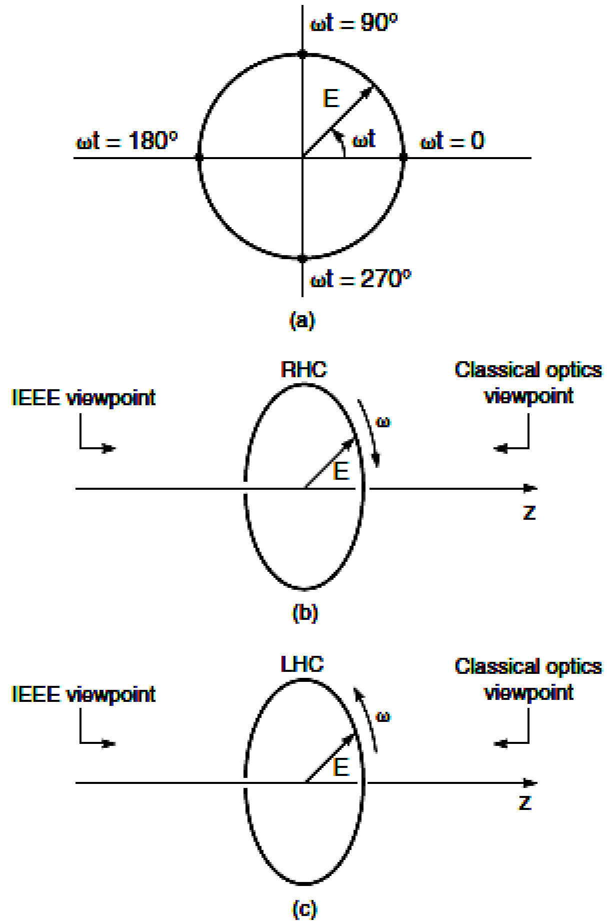


Figure 5.3: Circular Polarization

5.3 ANTENNA POLARIZATION

- It is defined by polarization (time-varying direction and amplitude) of the wave it transmits.

- The polarization of an antenna is the orientation of the electric field (E-plane) of the radio wave with respect to the Earth's surface and is determined by the physical structure of the antenna and by its orientation. It has nothing in common with antenna directionality terms: "horizontal", "vertical", and "circular". Thus, a simple straight wire antenna will have one polarization when mounted vertically, and a different polarization when mounted horizontally.

- Reflections generally influence polarization. For radio waves the most significant reflector is the ionosphere - signals which reflect from it will have their polarization changed unpredictably. For signals which are reflected by the ionosphere, polarization cannot be relied upon. For line-of-sight communications for which polarization can be relied upon, it can make a large difference in signal quality to have the transmitter and receiver using the same polarization; the difference which is commonly seen and this is more than enough to make the difference between reasonable communication and a broken link.

- Polarization is largely predictable from antenna construction but, especially in directional antennas, the polarization of side lobes can be quite different from that of the main propagation lobe. For radio antennas, polarization corresponds to the orientation of the radiating element in an antenna. A vertical Omni-directional Wi-Fi antenna will have vertical polarization (the most common type). An exception is a class of elongated waveguide antennas in which vertically placed antennas is horizontally polarized. Many commercial antennas are marked as to the polarization of their emitted signals.

- Horizontal dipole produces a horizontally polarized wave; vertical dipole produces a vertically polarized wave. When both these dipoles are mounted close to each other at right angles, they produce a circularly polarized wave.

- At the receiving end, antenna should be aligned in such a way that maximum power transfer should occur.
  - ✓ If the dipole is parallel to electric field E, induced voltage will be maximum; **V= max**
  - ✓ If dipole is perpendicular to E, then **V = 0**
  - ✓ If dipole lies in the plane of polarization and is at an angle α to E then **V = V max cos α**

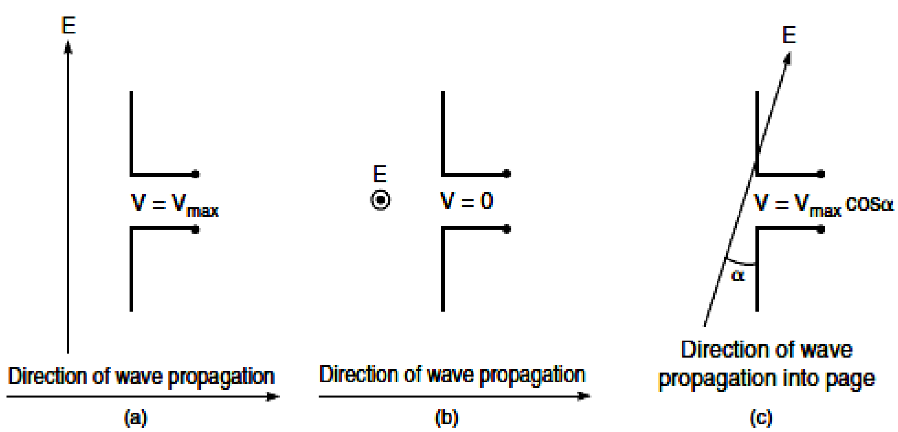


Figure 5.4: Linear polarization relative to a receiving dipole.

(If the wave is circularly polarized, maximum power is attained by the two dipoles producing it.)

- Grid of parallel wires is created to reflect a linear polarized wave when the electric field is parallel to the wires. This leads to a transmission of orthogonal wave.

- The simplest linear polarizer in concept is the wire-grid polarizer, which consists of a regular array of fine parallel metallic wires, placed in a plane perpendicular to the incident beam. Electromagnetic waves which have a component of their electric fields aligned parallel to the wires induce the movement of electrons along the length of the wires. As the electrons are free to move in this direction, the polarizer acts in a similar manner to the surface of a metal when reflecting light; and the wave is reflected backwards along the incident beam (minus a small amount of energy lost to joule heating of the wire).

- For waves with electric fields perpendicular to the wires, the electrons cannot move very far across the width of each wire; therefore, little energy is reflected, and the incident wave is able to pass through the grid. Since electric field components parallel to the wires are reflected, the transmitted wave has an electric field purely in the direction perpendicular to the wires, and is thus linearly polarized.

- For practical use, the separation distance between the wires must be less than the wavelength of the radiation, and the wire width should be a small fraction of this distance. This means that wire-grid polarizers are generally only used for microwaves and for far- and mid-infrared light. Using advanced lithographic techniques, very tight pitch metallic grids can be made which polarize visible light. Since the degree of polarization depends little on wavelength and angle of incidence, they are used for broad-band applications such as projection.

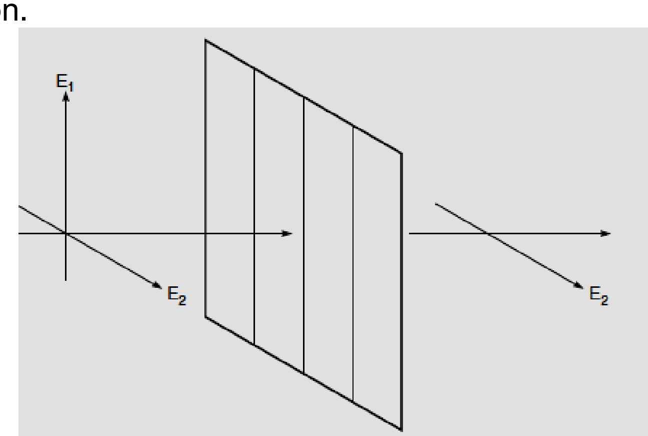


Figure 5.5: Wire Grid polarizer

5.4 POLARIZATION OF SATELLITE SIGNALS

- For a Geostationary Earth orbiting satellite transmitting a linear polarized wave, its horizontal polarization will be where the electric field is parallel to Earth's equatorial plane and its vertical polarization is where its Earth's electric field is parallel to the Earth's polar axis.

- For sub-satellite point on the equator, both the polarization vector will be at some angle relative to a reference plane. The reference plane for direction of propagation and local gravity direction is defined in the figure below.
-

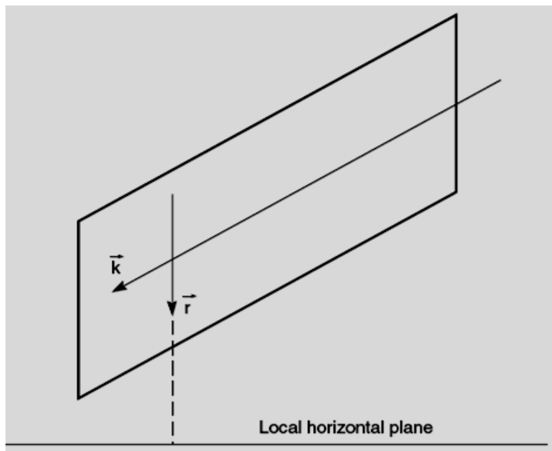


Figure 5.6: Reference plane for direction of propagation and local gravity direction

<u>Derivation:</u>

I) let k → propagation direction

r→ local gravity direction

f→ direction of normal to the reference plane.

f = k x r

p→ unit polarization vector at Earth Station

n→ Angle between p and f given by vector dot product

n = arcos (p.f / lfl)

As the angle between normal and its plane is 90^0, angle between p and its reference plane is given by:

ξ = |90^0 – nl

ξ = larcsin (p.f) / lfl l

Note: Polarization vector is always at right angles to the direction of propagation.

II) Relation between p and e

p→ polarization vector

e→ polarization at satellite

Vp→e lies parallel to Earth's North-South axis

Hp → e lies in the equatorial plane at right angles to the geostationary radius a_{gso} to the satellite.

The cross product is: **g = k x e**

g→ normal to the plane containing e and k

Cross product of g with k is given by h

$$\mathbf{h = g \times k}$$

Unit polarization vector at Earth is:

$$\mathbf{P = h / |h|}$$

III) Taking the longitude of satellite as reference, the satellite is positioned along the positive x axis at

$$\mathbf{x_s = a_{gso}}$$

Coordinates of Earth station position vector R are:

Rx → R cos λ cos B
Ry→ R cos λ sin B
Rz→ R sin λ

Where B = Φ_E - Φ_{SS} (Latitude of Earth Station and Sub-Satellite point)

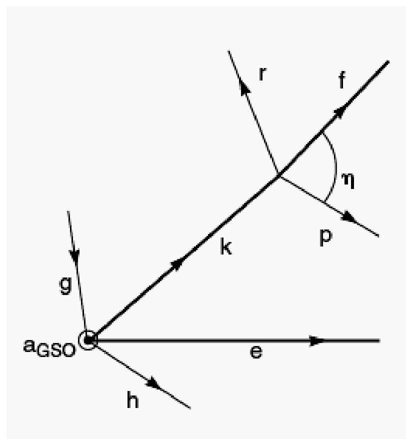


Figure 5.7: Vectors g = k x e and h = g x k

Local gravity direction is r = - R.
Thus the coordinates for direction of propagation are

$k_x = R_x - a_{gso}$
$k_y = R_y$
$k_z = R_z$

<u>Example:</u> A GEO satellite is positioned at 105^0 W and sends a vertically polarized wave. Determine its angle of polarization at an Earth station positioned 18^0 N latitude and 73^0 W longitudes.

Solution **:**

$$\lambda := 18 \cdot deg \qquad \phi_E := -73 \cdot deg \qquad \phi_{SS} := -105 \cdot deg$$

$$a_{GSO} := 42164 \cdot km \qquad R := 6371 \cdot km$$

(spherical earth of mean radius R assumed)

Calculations:

$$B := \phi_E - \phi_{SS} \qquad \text{...Eq. (3.8)}$$

Applying Eq. (5.15), the geocentric-equatorial coordinates for the earth station position vector are

$$\begin{bmatrix} R_x \\ R_y \\ R_z \end{bmatrix} := \begin{bmatrix} R \cdot \cos(\lambda) \cdot \cos(B) \\ R \cdot \cos(\lambda) \cdot \sin(B) \\ R \cdot \sin(\lambda) \end{bmatrix} \qquad \begin{bmatrix} R_x \\ R_y \\ R_z \end{bmatrix} = \begin{pmatrix} 5138.5 \\ 3210.9 \\ 1968.7 \end{pmatrix} \cdot km$$

The coordinates for the local gravity direction are

$$r := -\begin{bmatrix} R_x \\ R_y \\ R_z \end{bmatrix}$$

From Eq. (5.16), the geocentric-equatorial coordinates for the propagation direction are

$$k := \begin{bmatrix} R_x - a_{GSO} \\ R_y \\ R_z \end{bmatrix} \qquad k = \begin{bmatrix} -3.703 \cdot 10^7 \\ 3.211 \cdot 10^6 \\ 1.969 \cdot 10^6 \end{bmatrix} \cdot m$$

For vertical polarization at the satellite, the geocentric-equatorial coordinates for the polarization vector are $x = 0$, $y = 0$, and $z = 1$:

$$e := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$f := k \times r \qquad f = \begin{bmatrix} 0 \\ -8.3 \cdot 10^7 \\ 1.4 \cdot 10^8 \end{bmatrix} \cdot km^2$$

$$g := k \times e \qquad g = \begin{pmatrix} 3210.9 \\ 37025.5 \\ 0 \end{pmatrix} \cdot km$$

$$h := g \times k \qquad h = \begin{bmatrix} 7.3 \cdot 10^{7} \\ -6.3 \cdot 10^{6} \\ 1.4 \cdot 10^{9} \end{bmatrix} \cdot km^{2}$$

$$p := \frac{h}{|h|} \qquad p = \begin{pmatrix} 0.053 \\ -0.005 \\ 0.999 \end{pmatrix}$$

$$\xi := \operatorname{asin}\left(\frac{p \cdot f}{|f|}\right) \qquad \xi = 58.6 \cdot deg$$

= = = = = = = = = = = =

5.5 POLARIZATION DISCRIMINATION

- The propagation path of a signal has to go through the ionosphere in order to reach the satellite from earth station and vice versa. These signal sometimes also pass through clouds, rain crystals, rain etc. all these factors alter the polarization of a signal. Sometimes an additional orthogonal signal might get generated leading to the effect of depolarization. Two measures are used to calculate the effect of depolarization.

I) <u>Cross Polarization Discrimination (XPD)</u>

It's the most widely used measure. The transmitted electric field has magnitude E_1 before it enters the medium (as shown in figure 5.6). On entering the medium, E_1 also gets depolarized. Thus at the receiving end antenna, electric field may have two components:

Co-polar component with magnitude $\mathbf{E_{11}}$

Cross-polar component with magnitude $\mathbf{E_{12}}$

XPD in decibels is: $\mathbf{XPD = 20 \log E_{11} / E_{12}}$

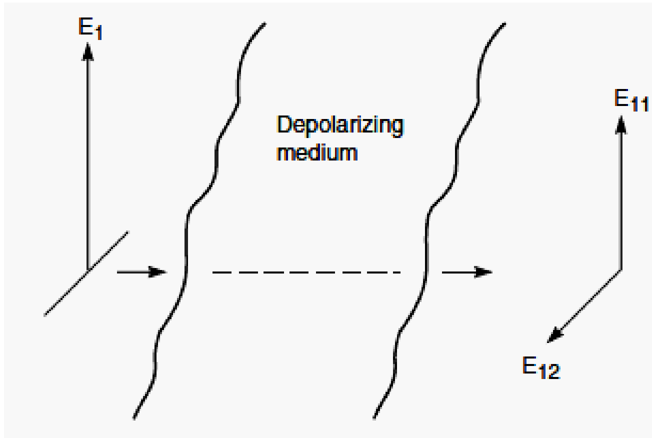


Figure 5.8: Vectors defining Cross Polarization Discrimation

II) **Polarization Isolation (I)**

Here the two orthogonally polarized signals with magnitudes E1 and E2 are transmitted. After passing through the depolarizing medium, the co-polar and cross-polar components exist for both the waves. The polarization isolation is defined by the ration of the received co-polar power to the received cross-polar power. This means it considers the additional depolarization.

$$I = 20 \log E_{11} / E_{12}$$

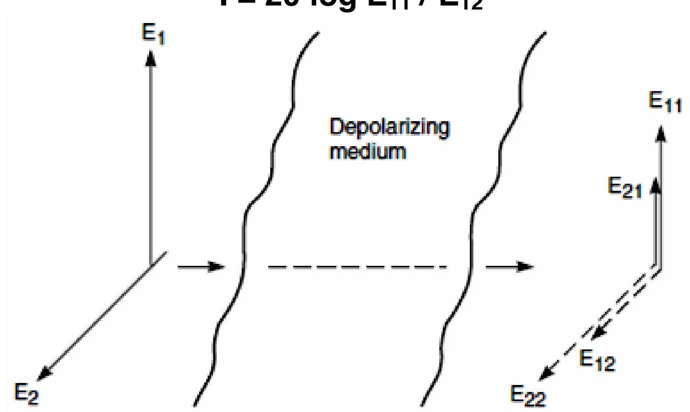


Figure 5.9: Vectors defining Polarization Isolation

- I and XPD have similar results when transmitted signals have the same magnitude ($E_1 = E_2$). The receiving system introduces negligible depolarization.

5.6 IONOSPHERIC DEPOLARIZATION

- The layer above the Earth's surface carrying charged particles is called ionosphere (as described in Unit 4). This layer is charged because of solar radiation.

- Free electrons that are freely (non-uniformly) distributed in this layer form electron clouds. And cause travelling ionospheric disturbances. Signals passing through these clouds face fluctuations and Faraday's rotation.

- The motion of these free electrons is within the magnetic field of the Earth. But as soon as the travelling wave passes through them, their direction of motion no longer remains parallel to the electric field of the wave and these electrons react back on these waves.

- This leads to a shift in the polarization. This angular shift in polarization is also dependent on the length of the path in the ionosphere and the electron density of the region of interest. Faraday's rotation is inversely proportional to the frequency shared and is not considered as a problem for frequency above 10 GHz.

  Let linear polarization wave have electric fields E at the receiver antenna with no Faraday rotation. Thus received power is E^2
  Faraday rotation of θ_F degrees results into co-polar component of the received signal being reduced to:

$$\mathbf{E_{co} = E \cos \theta_F}$$

  Thus the receiving power in this case is E_{co}^2
  Thus polarization loss (PL) in decibels is:

$$\mathbf{PL = 20 \log E_{co} / E}$$

$$\text{Thus, } \mathbf{PL = 20 \log (\cos \theta_F)}$$

  Similarly, cross-polar component is:

$$\mathbf{XPD = 20 \log E_{co} / E_x}$$

$$\text{Thus, } \mathbf{XPD = 20 \log (\cos \theta_F)}$$

5.7 RAIN DEPOLARIZATION

- The ideal shape of a rain drop is spherical as it minimizes the energy required to hold the raindrop together. Longer drops are oblate or elongated and they generally flatten underneath due to the gravity or air resistance.

- Linear polarized wave can have two components; they are vertical and horizontal polarized. Considering a wave with its electric vector at some angle τ relative to the major axis of the rain drop.

- Vertical component is parallel to the minor axis of the rain drop and thus having less water content. It is vice versa for the horizontal component. Due to this, there lies a difference between the attenuation and phase shift of each electric field component.
- This is termed as differential attenuation and differentia phase shift and they result in depolarization of the signal.

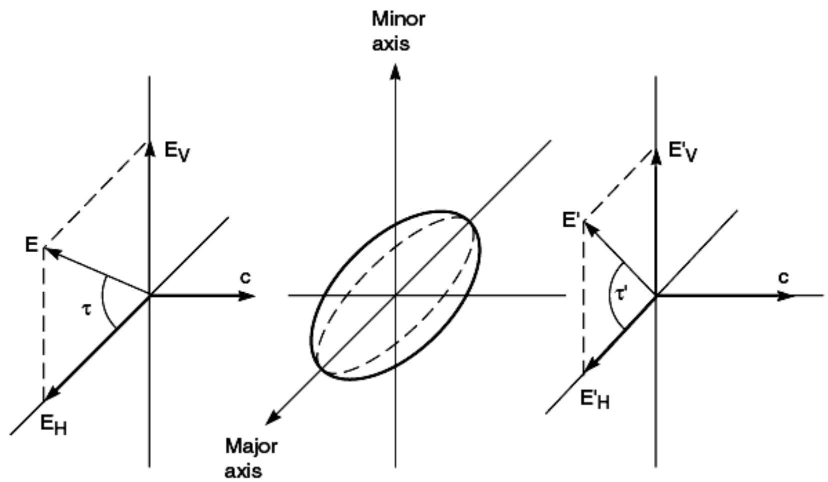


Figure 5.10: Polarization vector relative to major and minor axis of a rain drop.

- Compensation techniques for rain depolarization in satellite links are discussed from the viewpoint of system configuration, performance, and technical feasibility. The compensator using two rotatable polarizer has sufficient performance at microwave region. Two types of compensators, as combination of 90^0 and 180^0polarizer and a combination of two 90 deg polarizer, were developed and tested. Satisfactory results were obtained at 4-GHz band by using the Intelsat satellite.

- To summarize, the effect of rain depolarization has been calculated as a function of rain rate, path length, and frequency of differential phase shift and differential attenuation due to oblate raindrops. The experimental values generally agree with the theoretical predictions, but the latter tend to underestimate the effect. This underestimation is possibly due to some near-field effect plus the errors introduced by approximating the raindrops as ellipsoids, especially at high rain rates. Nevertheless, theoretical and experimental values of the level of depolarization due to rain generally differ by no more than 3dB.

5.8 ICE DEPOLARIZATION

- Ice layer that is present over the rain region causes depolarization. Ice which is comparatively a good dielectric component leads to more loss. Shape of ice crystals is like a needle or like a plate and if they are all aligned; their effect of depolarization is much more.
- Sudden increase in the XPD coinciding with lightening can cause severe damage to the signals. Generally a fixed value in decibels of XPD is calculated for account of ice depolarization. [***Dielectric***: it is an electric insulator that can be polarizes by an applied electric field.]

5.9 SUMMARY

Polarization is an important factor for satellite communication. Both RF antennas and electromagnetic waves are said to have a polarization. For the electromagnetic wave the polarization is effectively the plane in which the electric wave vibrates. This is important when looking at antennas because they are sensitive to polarization, and generally only receive or transmit a signal with a particular polarization.

For most antennas it is very easy to determine the polarization. It is simply in the same plane as the elements of the antenna. So a vertical antenna (i.e. one with vertical elements) will receive vertically polarized signals best and similarly a horizontal antenna will receive horizontally polarized signals.

5.10 EXERCISE

1. Discuss the plane TEM wave.
2. Write a note on Ice Depolarization.
3. Differentiate between XPD and Polarization Isolation.
4. What is Polarization Discrimination?
5. What are travelling ionospheric disturbances?
6. Write a note on Rain Depolarization.
7. Explain what is meant by orthogonal polarization and the importance of this in satellite communication.
8. Discuss Ionospheric Depolarization.

References

33. www.wikipedia.com
34. http://www.tech-faq.com/vsat.html
35. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
36. Rappaort, Wireless Communications Principals and Practices
37. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
38. P. Nicopolitidis ,Wireless Networks, John Wiley
39. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
40. http://www.britannica.com/EBchecked/topic/524891/satellite-communication

❖❖❖❖

6A

ANTENNAS

Contents:

6A.1 INTRODUCTION

- An antenna is used to radiate electromagnetic energy efficiently and in desired directions. Antennas act as matching systems between sources of electromagnetic energy and space. The goal in using antennas is to optimize this matching. Here is a list of some of the properties of antennas:
  1) Field intensity for various directions (antenna pattern).
  2) Total power radiated when the antenna is excited by a current or voltage of known intensity (or Power Flux Density)
  3) Radiation efficiency which is the ratio of power radiated to the total power (Radiation Pattern).
  4) The input impedance of the antenna for maximum power transfer (matching).
  5) The bandwidth of the antenna or range of frequencies over which the above properties are nearly constant.

- Antennas can also be classified as electrical devices which convert electric currents into radio waves and vice-versa. They are generally used with a radio transmitter and receiver. They are broadly classified in two categories:

- Transmitting antennas
- Receiving antennas

Difference is in the mode of operation, different functions etc. as the transmitting as well as the receiving antenna.

- Earth Station Antenna
- Satellite Antenna

Difference is mainly in their environmental conditions which lead to their different designs.

- Typically an antenna has an array of metallic conductors that are electrically connected. An oscillating current of electrons focused through the antenna by a transmitter creates an oscillating electric field.

- These fields are time-varying and radiate from the antenna into the space as a moving electromagnetic field wave. Certain properties of antennas such as directional characters result into reciprocity theorem.

- Different types of antennas are:
  1) ***Dipole Antennas***: The dipole is one of the most common antennas. It consists of a straight conductor excited by a voltage from a transmission line or a waveguide. Dipoles are easy to make.

  2) ***Aperture Antennas***: A horn as shown in the figure below is an example of an aperture antenna. These types of antennas are used in Satellite spacecraft more commonly.

  3) ***Reflector Antennas***: The parabolic reflector is a good example of reflectors at microwave frequencies. In the past, parabolic reflectors were used mainly in space applications but today they are very popular and are used by almost everyone who wishes to receive the large number of television channels transmitted all over the globe.

  4) ***Array Antennas:*** A grouping of similar or different antennas form an array antenna. The control of phase shift from element to element is used to scan electronically the direction of radiation.

  5) ***Loop Antennas:*** A loop of wire, with many turns, is used to radiate or receive electromagnetic energy creates a loop antenna. These antennas can be used at home to capture signals of World radios or local television channels.

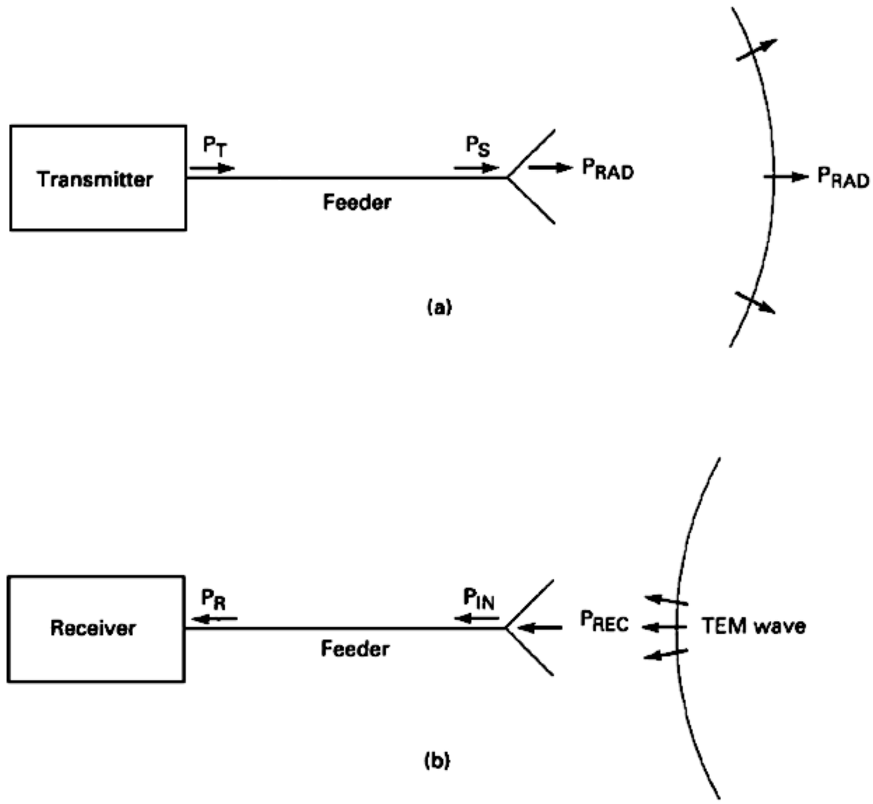


Figure 6.1: a) Transmitting Antenna; b) Receiving Antenna

a) Power amplifier is transmitter as P_T watts. Feeder connects to this antenna and the new power reaching to this antenna will be P_T – Losses at the feeder. Further loss is seen in the antenna and thus the radiated power is shown as P_{RAD}.

b) Power P_{Rec} is transferred to the antenna from a passing radio wave. Again the losses in the antenna will reduce the power at the feeder. Giving only P_{IN} to the feeder. Receiving feeder looses further reduce the power to P_R.

Definitions:

EMF (Electromagnetic Force): The electromagnetic force is a special force that affects everything in the universe because (like gravity) it has an infinite range. It has the ability to attract and repel charges. Since material in solid and liquid forms are made of charges having a unique order, they, too, may be manipulated by this force. It is also responsible for giving things strength, shape, and hardness. The electromagnetic force can be generated by three types of fields known as the electrostatic field, magneto-static field, and the electromagnetic field.

Gain: An antenna's power gain or simply gain is a key performance figure which combines the antenna's directivity and electrical efficiency. For a transmitting antenna gain is how well the antenna converts input power into radio waves headed in a specified direction. For a receiving antenna gain is how well the antenna converts radio waves arriving from a specified direction into electrical power. When no direction is specified, "gain" is understood to refer to the peak value of the gain. A plot of the gain as a function of direction is called the radiation pattern.

6A.2 RECIPROCITY THEOREM

- The Reciprocity Theorem states that if current *I* is induced in an Antenna B which is working on receiving mode, this current is applied by the EMF at the terminals of the antenna A that is working on the transmitting mode. Now if the same EMF is applied to the terminals of B, then it will induce the same current at the terminals of A.

- There are a number of consequences of this theorem:
    - In practical use, it is seen that antennas transmit more energy in some directions than others; and similarly, they also receive more energy while pointing in some directions than others.
    - Thus the directional pattern for an antenna operating in the transmit mode is the same as that when operating in the receive mode
    - Antenna’s impedance is the same or both the modes of operation.

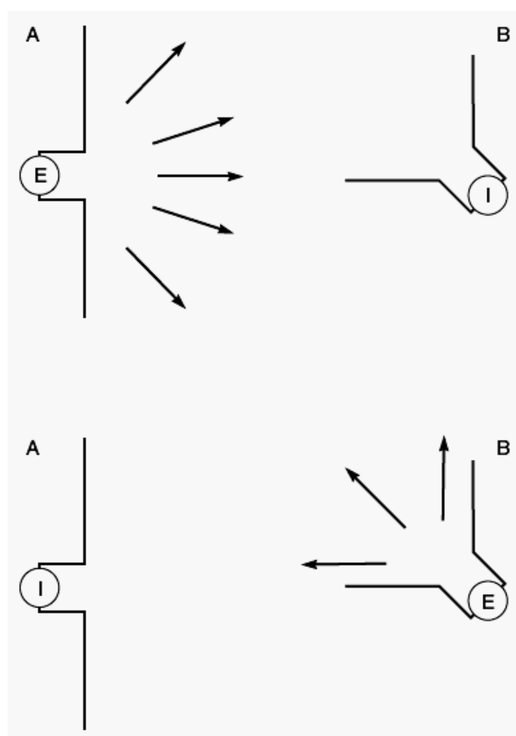


Figure 6.2: The reciprocity theorem

6A.3 POWER FLUX DENSITY

- It is the quantity used in calculating the performance of satellite communication links. The concept can be understood by:
    - Imagine the transmitting antenna is at the centre of sphere.
    - The power from the antenna radiated outward normal to the surface of the sphere; here the Power Flux Density is the power flow per unit surface area.
    - The Power Flux Density is a vector quantity given by:

$$\Psi = E^2 / Zw$$

(Here, E is the resultant magnitude of electric field in volts and Zw is in Ohms.)

6A.4 ISOTROPIC RADIATOR AND ANTENNA GAIN

- Isotropic radiator means radiating energy equally in all directions. This is a theoretical concept as no antenna (irrespective of where it is placed) can radiate signals equally in all directions.

- This concept is used for comparison purpose with the real antennas. As proved by the Reciprocity Theorem, $P_{RAD} = P_{S.}$ Let the Isotropic Radiation be at the centre of radius r, the Power Flux Density per unit area is:

$$\Psi = P_S / 4 \pi r^2$$

- As Flux Density forms a real antenna will change with direction and its maximum will occur, the gain of this antenna id the ratio of this maximum to that for the isotropic radiator at the same radius r.

$$G = \Psi_M / \Psi_i$$

- Gain is related to directivity. To improve the directivity, the difference between the power radiated by the real antenna and the power to it needs to be considered.

$$P_{RAD} = \eta_A P_S$$

- Where η_A is the antenna efficiency. The Gain by the antenna is called isotropic gain denoted by Gi. Gain varies from different types of antennas.

6A.5 COORDINATE SYSTEM

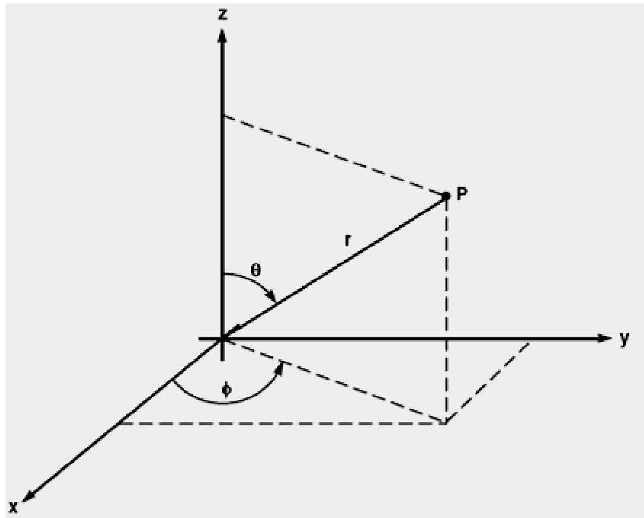


Figure 6.3: Spherical Co-ordinate system

- The antennas are located on the surface and rotate with respect to a source in the sky due the rotation of the earth. For aperture separation the antenna positions are specified in a co-ordinate system such that the separation of the antennas is the projected separation in plane normal to the phase centre. In such a co-ordinate system the separation between the antennas is as seen by the observer sitting in the source reference frame.

- An antenna coordinate system is implicit in almost every antenna measurement. Terms such as pattern, main and cross-component, beam pointing and peak gain, imply the definition of directions and/or vector field components, which require a coordinate system. Reference is often made to the cross-component of an antenna as if there is a unique definition of such a quantity when in fact the cross-component as well as the main component will depend on which coordinate system is used and how it is oriented with respect to the antenna.

- A spherical coordinate system is a coordinate system for three-dimensional space where the position of a point is specified by three numbers: the radial distance of that point from a fixed origin, its inclination angle measured from a fixed direction, and the azimuth angle of its orthogonal projection on a reference plane that passes through the origin and is orthogonal to the zenith, measured from a fixed reference direction on that plane. The inclination angle is often replaced by the elevation angle measured from the reference plane.

- In order to discuss the directional patterns of an antenna, it is necessary to set up the coordinate system to which these can be referred. With reference to figure 6.3, the antenna is imagined to be at the origin of the coordinates, and at a distant point in space is related to the origin by the coordinates r, θ and Φ. Thus r is the radius vector, the magnitude of which gives the distance between point P and the antenna; Φ is the angle measured from the x axis to the projection of r in the x plane; and θ is the angle measures from the z axis to r.

- The x, y and z axes form a right-hand set. When one looks at the positive z direction, a clock-wise rotation is required to move from positive x axis to the positive y axis. This becomes particularly significant when the polarization of the radio waves associated with antennas is described.

6A.6 RADIATION PATTERN

- Radiation pattern shows that the gain of an antenna varies with direction. At a fixed distance r, gain will vary with θ and Φ and is written as G (θ, Φ) as shown in figure 6.3. The Radiation Pattern is the gain normalized at a maximum value.

- Let the maximum gain be G. thus Radiation Pattern is denoted as g is given by:

$$g(\theta, \Phi) = G(\theta, \Phi) / G$$

- The Radiation Pattern gives directional properties of the antenna normalized to the maximum value (maximum gain). A three-dimensional representation of this shows a lobe.

- In this figure 6.4, the length of the radius line to any point on the surface of the lobe gives the value of radiation function at that point. The minimum lobe shows the beam of radiation and the beamwidth id shown as an angle by three-dimensional lines.

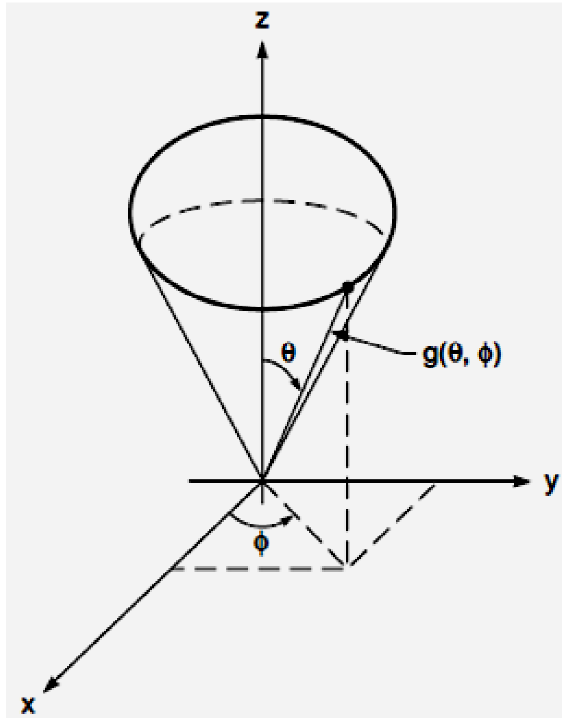
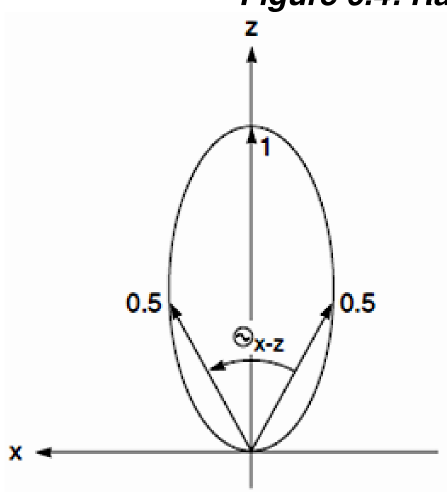


Figure 6.4: Radiation Pattern



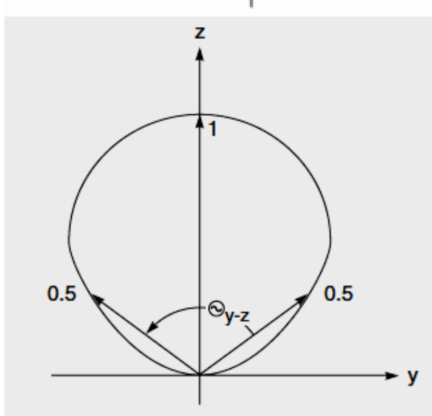


Figure 6.5: Radiation Pattern of H-Plane *Figure 6.6: Radiation Pattern of E-Plane*

- Generally, a beamwidth may be asymmetrical, then beamwidth in the H-plane is $\Phi = 0^0$ as shown in figure 6.5 and in E plane $\Phi = 90^0$ is seen in figure 6.6.

6A.6 ANTENNA APERTURE

- Antenna aperture can be visualized as the area of a circle constructed broadside to incoming radiation where the entire radiation passing within the circle is delivered by the antenna. Antenna gain can be increased by directing radiation in a single direction and thus reducing it in all the other directions.

- Hence layered antenna apertures can produce a higher gain and narrower beamwidth. The most common aperture antenna is Horn Antenna and Reflector Antenna.

- Figure 6.6 shows an idealized aperture. It consists of a rectangular aperture of sites a and b cut in an infinite ground plane. Here radiation form some parts adds constructively in some direction and destructively in others. This in result will exhibit a main lobe and a number of side lobes.

- The actual distribution of energy depends on the manner in which the aperture is energized. Distribution pattern also depends upon the physical construction of the antenna.

- As cross-polarization can occur with real antennas an unwanted signal is transmitted along with the signal send and received along with the incoming data. The unwanted signal causes interference. Hence this aspect has to be taken care of along with other practical consideration.

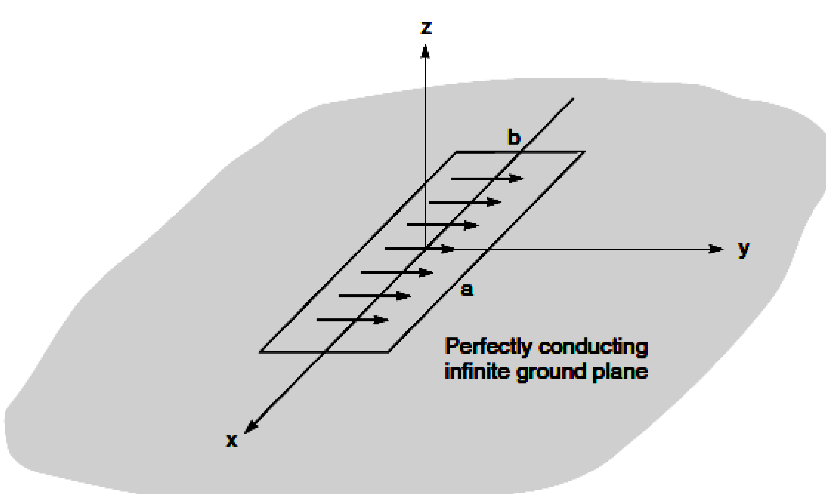


Figure 6.7: An idealized Aperture Radiator

6A.7 SUMMARY

- This unit describes the concept of antennas. An antenna is a device for converting electromagnetic radiation in space into electrical currents in conductors or vice-versa, depending on whether it is being used for receiving or for transmitting, respectively. It further gives the basic categories of antennas used in satellite communication. Further, this unit discusses the Reciprocity theorem which states that: The reciprocity theorem states that if an emf E in one branch of a reciprocal network produces a current I in another, then if the emf E is moved from the first to the second branch, it will cause the same current in the first branch.

  After defining the reciprocity theorem for antennas, this unit describes the other measures needed to calculate the functionality, position of an antenna. They are:

Power Flux Density: It calculates the performance of a communication link.

Isotropic Radiator: Defines the theoretical concept of an ideal antenna.

Antenna Gain: It calculates the efficiency of an antenna.

Coordinate System: it is a coordinate system for three-dimensional space where the position of a point is specified.

Radiation Pattern: It elaborates on the concept that the gain of an antenna varies with direction and is different than what the theoretical concept describes.

Antenna Aperture: It is an area of a circle constructed broadside to incoming radiation where the entire radiation passing within the circle is delivered by the antenna.

6A.8 EXERCISE

1. Explain the concept that 'gain of an antenna varied with direction' using radiation pattern.
2. What is an isotropic radiator?
3. Explain what is meant by reciprocity theorem as applied to antennas?
4. What is an offset feed of an antenna?
5. What are Aperture Antennas?
6. Define power flux density.
7. How are antennas classified? With an appropriate diagram, show an antenna in the transmitting and receiving modes.

References

41. www.wikipedia.com
42. http://www.tech-faq.com/vsat.html
43. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
44. Rappaort, Wireless Communications Principals and Practices
45. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
46. P. Nicopolitidis ,Wireless Networks, John Wiley
47. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
48. http://www.britannica.com/EBchecked/topic/524891/satellite-communication

❖❖❖❖

6B

ANTENNAS

Contents:

6b.1 INTRODUCTION

- In the previous unit, we have already defined an antenna and the basic parameters needed to calculate the efficiency in functionality and position of a satellite. We have defined a few antennas like Different types of antennas are: *Dipole Antennas, Aperture Antennas*, *and Reflector Antennas etc*. Now we will see antennas which are actually used by the spacecraft or the earth station for satellite communication. They include the Horn antennas, parabolic reflectors which include Gregorian and Cassegrain antennas, double reflector antennas.

Definitions:

Waveguide: A waveguide is an arrangement which guides waves, such as electromagnetic waves or sound waves. There are different types of waveguides for every type of wave. The original meaning is a hollow conductive metal pipe used to carry high frequency radio waves, mainly microwaves. Waveguides differ in their geometry which can confine energy in one dimension such as in slab waveguides or two dimensions as in fibre or channel waveguides.

Antenna Feeder: antenna feed refers to the components of an antenna which feed the radio waves to the rest of the antenna structure, or in receiving antennas collect the incoming radio waves, convert them to electric currents and transmit them to the receiver. Antennas typically consist of a feed and additional reflecting or directive structures (such as a parabolic dish or parasitic elements) whose function is to form the radio waves from the feed into a beam or other desired radiation pattern. In simple antennas the feed usually consists of the feed antenna (driven element), the part of the antenna which actually converts the radio frequency currents to radio waves or vice versa, and the feed line (transmission line), which connects the feed antenna with the receiver or transmitter.

Hyperboloid: a hyperboloid is a quadric – a type of surface in three dimensions – described by the equation a2/x2 + b2/y2 – c2/z2 = 1.

6B.2 HORN ANTENNA

[One of the first horn antennas was constructed in 1897 by Indian radio researcher Jagadish Chandra Bose in his pioneering experiments with microwaves.]

- A horn antenna is an example of aperture antenna [refer unit VI-a] which provides a smooth transition from a waveguide to a larger aperture that couples more effectively into the space. In these antennas a flaring metal waveguide is shaped like a horn to direct radio waves in a beam.

- These antennas are designed to cover large areas on Earth. They are also used as primary feeds for reflector type antennas in transmitting as well as receiving modes. Three common types of horn antennas used for satellite communication are given below.

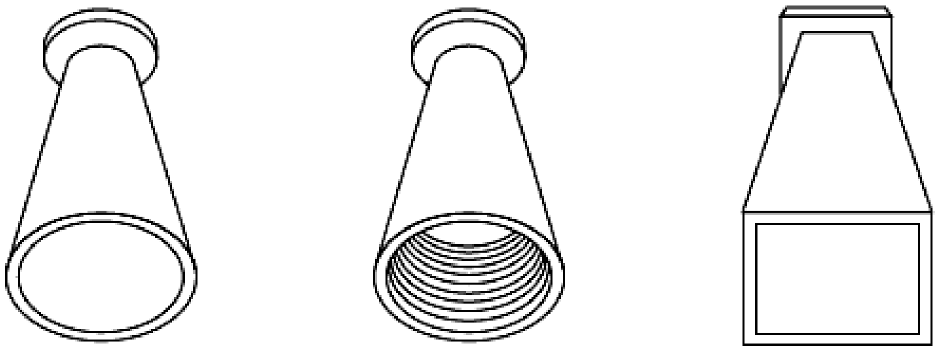


Figure 6b.1: Horn Antennas- a) Smooth-Walled Conical b) Corrugate c) Pyramidal

6b.2.1) Conical Horn Antenna

- This antenna is shown in figure 6b.1 (a). It is the simplest form of horn antenna. The inner wall of this antenna is smooth. These antennas can accept signals in rectangular form and converts them into circular wave using rectangular-to-circular transition device.

- Horn antennas may be used with circular or linear polarized wave. The smooth-walled horn antenna does not produce a symmetrical main beam even though the horn is symmetrical. This lack of symmetry is a disadvantage where global coverage is required.

  Electrical Field Distribution of Horn Antennas: The curved lines can resolve into horizontal and vertical components. Transverse electromagnetic wave is linearly polarized. Horizontal component of aperture field gives rise to the cross-polarized waves in the far-field region. As shown in figure 6b.2.

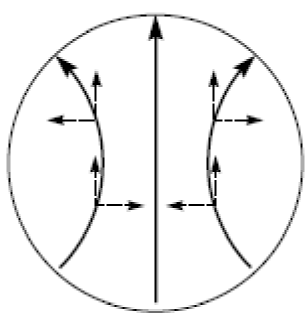


Figure 6b.2: Aperture field in smooth-walled conical horn

6b.2.2) Corrugate Horn Antenna

- By making conical horn antenna work on non-linear combination of transverse electric and transverse magnetic modes, the cross-polarization can be reduced and more efficient beam is produced.

- This can be done by making the inside wall of the horn antenna corrugate, thus making a new class of antenna called as Corrugate Horn antenna.

6b.2.3) Pyramidal Horn Antenna

- These antennas are designed for linear polarization. Its rectangular cross section a x b operates in the Transverse Electromagnetic wave mode.

- Beam width of these antennas differs in E and H planes. It can operate on the H and V polarized mode simultaneously giving rise to dual-linear polarization.

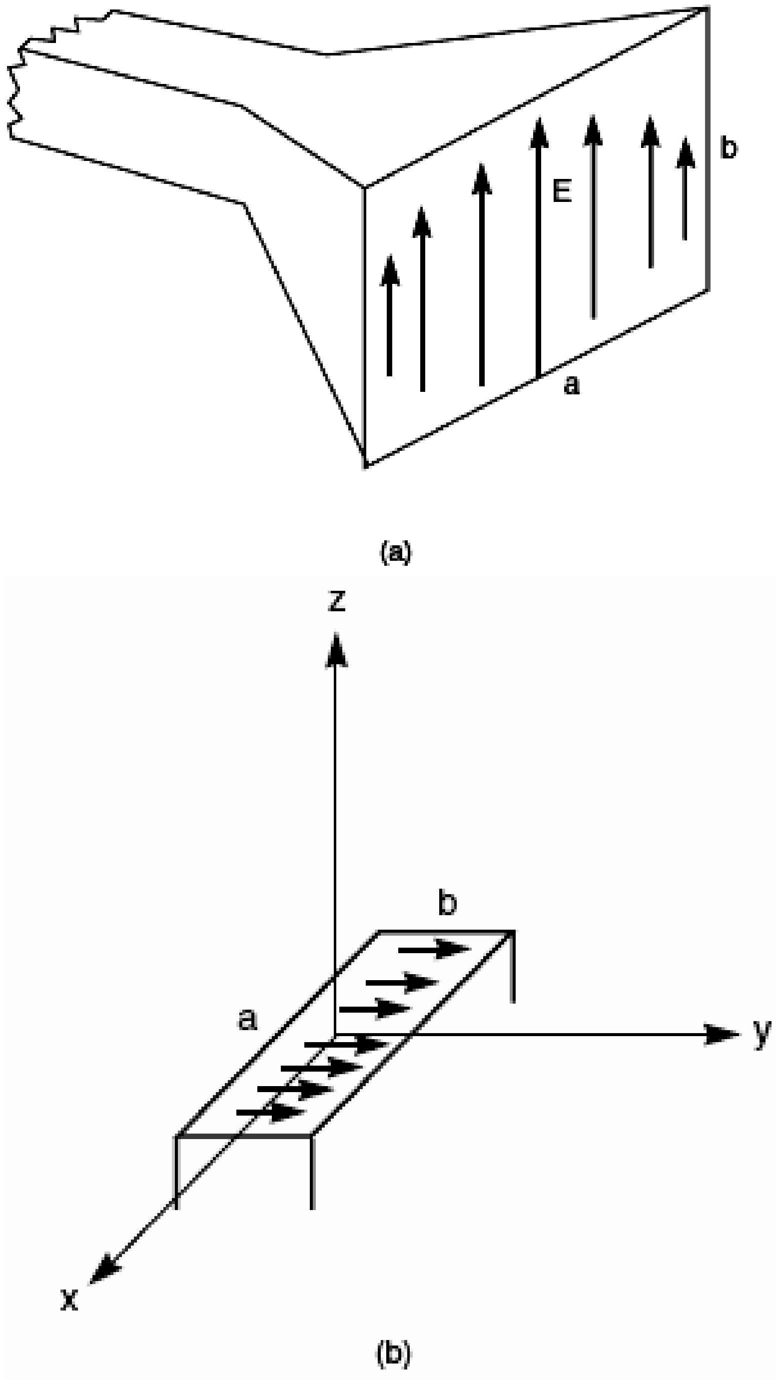


Figure 6b.3: The Pyramidal horn

6B.3 PARABOLIC REFLECTOR

- They are used to increase the gain of antennas. A parabolic reflector has a focusing device which helps to concentrate in one direction. Example: used in homes for reception of TV signals. The focusing property is associated with light.

- Parallel rays of light falling on the reflector get converged on a single point called focus. Similarly, rays originating from the focus are reflected as a parallel beam of light. This property of light is applicable to electromagnetic waves.

- The parabolic reflector, which takes the shape of a parabola, traces out a curve with the reflector on any plane normal to the aperture plane containing the focus. Radiations from these antennas appear to originate as a plane wave from a plane normal to the axis and containing a directrix (directrix is a fixed line used in describing a curve or surface).
  - The radio link behaves like a far-field component.
  - Reflected wave is a plane wave.
  - Waves falling on the reflector, originating from an isotropic source have a spherical wavefront.
  - The power flux density is a plane wave and is independent of the distance.
  - Power flux density in a plane wave decreases in a far-field component and is inverse to the distance squared.

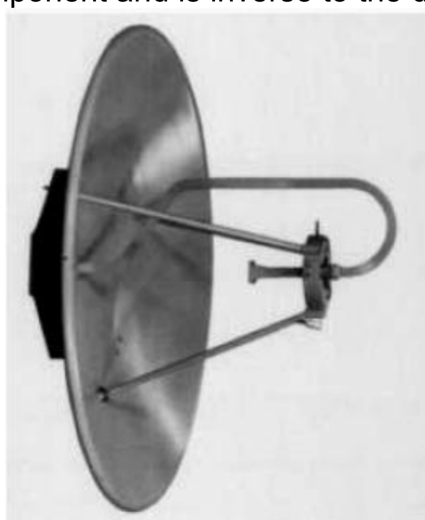


Figure 6b.4: a Parabolic Reflector

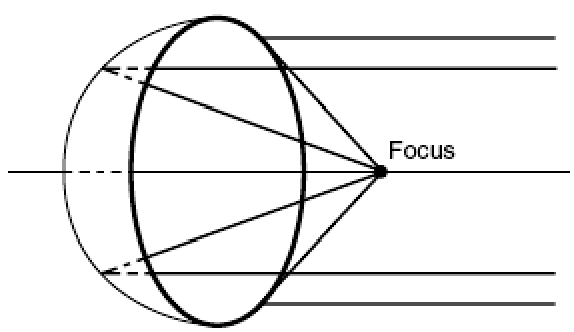


Figure 6b.5: Parabolic Reflector's focusing property

6B.4 OFFSET FEED

- Figure 6b.6 shows a paraboloidal reflector with horn feed at the focus. In this instance, the radiation pattern of the horn is offset so that it illuminates only the upper portion of the reflector. He feed horn and the support can be placed well clear of the main beam so that no blockage occurs. Blockage results in 10 percent of the efficiency.

- The main disadvantage of offset feed are that a stronger mechanical support is required to maintain the reflector shape and because of the symmetry, the cross-polarization with a linear polarized feed is worst compared with the centre-fed antenna. Because of the limited transmitter power provided by their solar cells, satellite antennas must function as efficiently as possible.

- Polarization compensation can be introduced into the primary feed to correct for the cross-polarization or a polarization-purification grid can be incorporated into the antenna structure. Offset feed is also used with double-reflector earth station antennas and id being used increasingly with small receive-only earth station antennas. The offset design is also widely used in radar antennas. These must collect as much signal as possible in order to detect faint return signals from faraway targets.

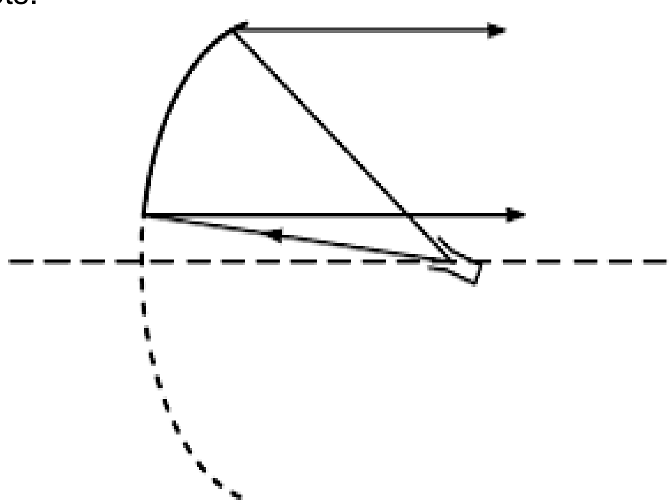


Figure 6b.6: Ray paths of a offset reflector

6B.5 DOUBLE-REFLECTOR ANTENNAS

- In these antennas, a short feeder is used to minimize losses. It is used by large earth stations where transmit power is sufficiently high and thus low receiver noise is required. From figure 6b.7; the rear mount makes a compact feed, which is an advantage where steerable antennas must be used and access for servicing becomes easier. The sub-reflector, which is mounted at the front of the main reflector, is generally smaller than he feed horn and causes less blockage.

- The two main types of double-reflector antennas used in satellite communication are:
  a) Cassegrain Antenna
  b) Gregorian Antenna

Figure 6b.7: Double Reflector Cassegrain Antenna.

6b.5.1) Cassegrain Antennas

- A Cassegrain antenna is a parabolic antenna in which the feed radiator is mounted at or behind the surface of the concave main parabolic reflector dish and is aimed at a smaller convex secondary reflector suspended in front of the primary reflector. The beam of radio waves from the feed illuminates the secondary reflector, which reflects it back to the main reflector dish, which reflects it forward again to form the desired beam.

- This antenna has a main parabolic and sub-reflector which is hyperboloid. The sub-reflector has two focal points; one focal point is made to coincide with that of the main reflector and the other with the phase centre of the feed horn.

- Uniform illumination is achieved in Cassegrain and these antennas are used by large earth stations.

- If the curvature of the centre section of this antenna is altered and is made greater than that the hyperboloid allows it to reflect more energy towards the edge of the main reflector, making the amplitude distribution more uniform.

- The curvature of the centre section is made smaller to compensate on the reduced path length so that the constant phase condition across the aperture is maintained. The edge of the sub-reflector is designed to reduce the “spill over”.

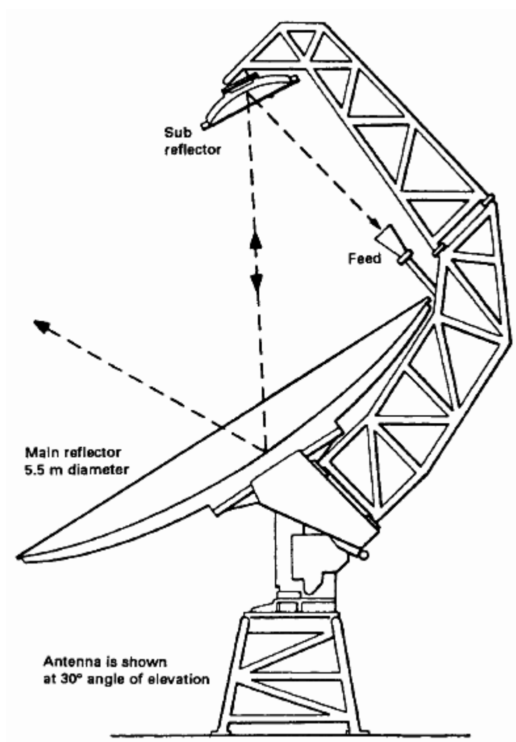


Figure 6b.7: Double Reflector Gregorian Antenna.

6b.5.1) Gregorian Antennas

- Gregorian antenna, like Cassegrain antenna is a parabolic reflector. It consists of a main paraboloid sub-reflector which is an ellipsoid with the hyperboloid. Sub-reflector has two focal points, one which is made to coincide with that of the main reflector and other with phase centre of the feed horn.

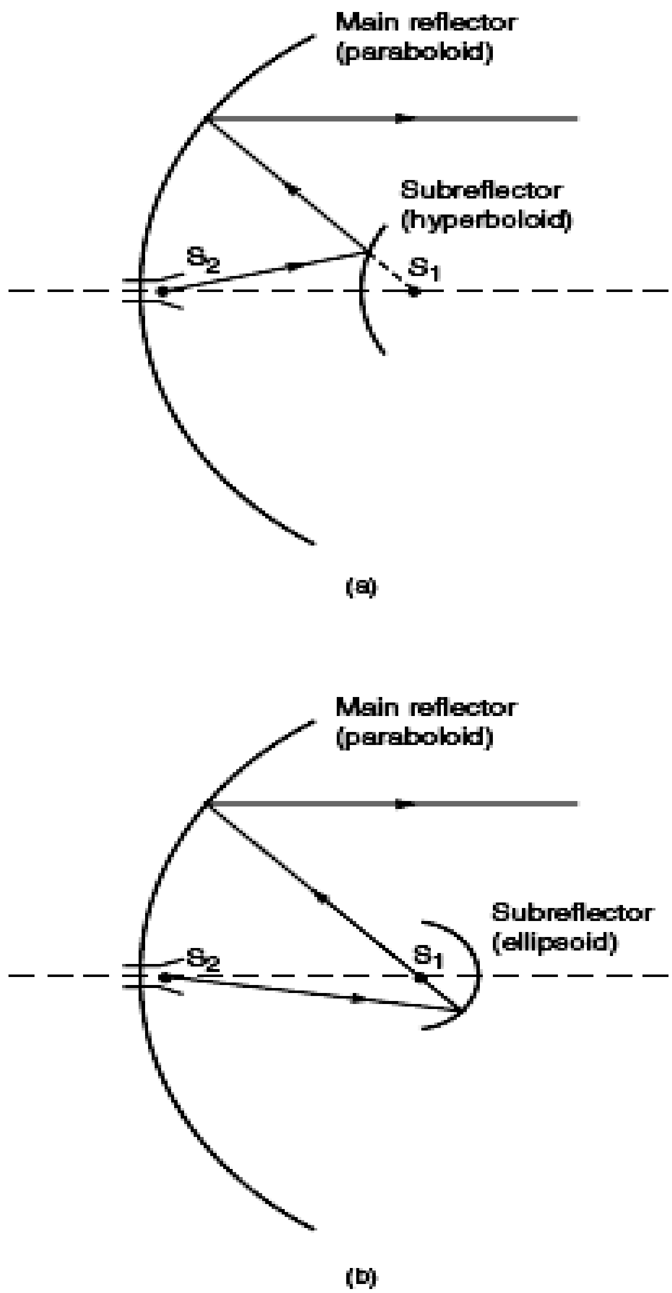


Figure 6b.8: Ray Paths for a) Cassegrain b) Gregorian Antennas.

6B.6 SHAPED REFLECTOR SYSTEM

- Ripples/ Dimples are created on the surface of the antenna. Their depth is approximately equal to the wavelength of the signal. Due to these ripples, reflection of the wave takes place. These reflections reinforce more radiation in some direction and less in others.

- This Shaped Reflector is also called Hughes Shaped Reflector. (Hughes Aircraft Company was a major American aerospace and defence contractor founded in 1932 by Howard Hughes in Culver City, California as a division of Hughes Tool Company. The company was known for producing, among other products, the Hughes H-4 Hercules "Spruce Goose" aircraft, the atmospheric entry probe carried by the Galileo spacecraft, and the AIM-4 Falcon guided missile.)

- ***Design of Shaped Reflector System:***
  - Design of these reflectors is based on the ground area of interest (where the antenna is to be places). A grid is lain on the map (again, the area of interest) and at each grid intersection a weighing factor is assigned which corresponds to antenna gain desired in that direction. Intersection points coincide with the Azimuth angle and Elevation angles (Refer Unit 1-2 for definition) of earth station.

  - The beam shaping stage starts by selecting a smooth parabolic reflector that forms an elliptical beam encompassing the coverage area.

  - The reflector surface is computer modeled as a series of mathematical functions which change until model produces a desired coverage.

- ***Performance Check of Shaped Reflector System:***
  - On the first pass, the computer analyses the function change and translates it into surface ripple.

  - The bam foot print computed for the ripple surface is compared with the coverage area.

  - The function change analysis is refined and the passes are repeated until a good match is obtained.

- ***Uses of Shaped Reflector System***
  - Shaped reflector systems are used to compensate for rainfall attenuation and are mostly used by direct broadcast satellites.

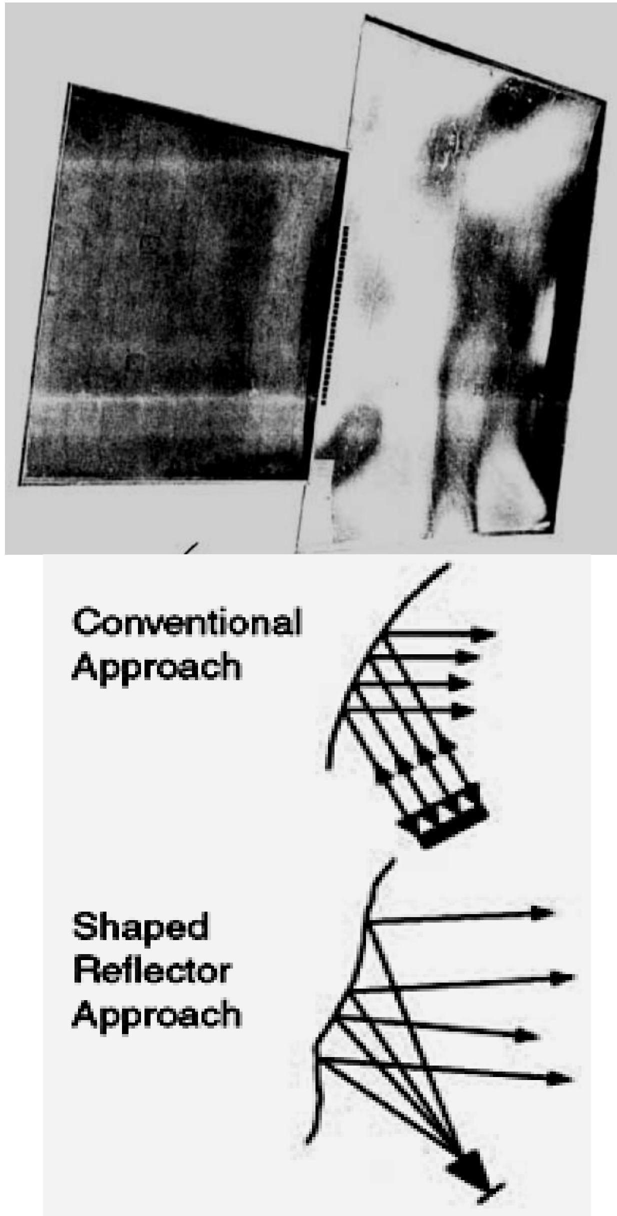


Figure 6b.9: Shaped- Reflector showing ray path.

6B.7 SUMMARY

In continuation with Unit Via, this unit discusses various antennas and their structures which are used by different types of satellites revolving around the earth for various purposes. This unit begins with the Horn Antennas, they are of three types: Conical, Corrugated and Pyramidal.

Further this unit discusses the Parabolic reflectors reflector which has a focusing device which helps to concentrate in one direction. Going further, the next topic discussed is the double-shaped reflector antennas. These antennas are generally located at the earth station and their structure resembles a parabola. These antennas are of two types depending on their shape; Cassegrain and Gregorian.

To conclude with, this unit discusses a discovery by Sir Hughes which is called as Shaped-Reflector system which is designed to trap the signal coming to the antenna with a special modification in antenna's shape and creating ripples on its surface for making the surface appropriate for reflection in certain angles only.

6B.8 EXERCISE

8. What is the need of creating ripples on the surface of Hughes shaped reflector antennas?
9. Write a note on describing the shaped reflector system.
10. What are double reflector antennas? Elaborate on its two types.
11. Write a note on Cassegrain Antenna.
12. Write a note on Gregorian Antenna.
13. What is an offset feed of an antenna?
14. How can parabolic reflectors used in satellite communication to enhance the gain of antennas?
15. Write a note on Horn antennas.

References

50. www.wikipedia.com
51. http://www.tech-faq.com/vsat.html
52. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
53. Rappaort, Wireless Communications Principals and Practices
54. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
55. P. Nicopolitidis ,Wireless Networks, John Wiley
56. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
57. http://www.britannica.com/EBchecked/topic/524891/satellite-communication

❖❖❖❖

7

COMMUNICATION SATELLITES PART I

Contents:

7.1 INTRODUCTION

- Communication satellites are specifically made for telecommunication purpose. They are used for mobile applications such as communication to ships, vehicles, planes, hand-held terminals and for TV and radio broadcasting.

- They are responsible for providing these services to an assigned region (area) on the earth. The power and bandwidth of these satellites depend upon the preferred size of the footprint, complexity of the traffic control protocol schemes and the cost of ground stations.

- A satellite works most efficiently when the transmissions are focused with a desired area. When the area is focused, then the emissions don't go outside that designated area and thus minimizing the interference to the other systems. This leads more efficient spectrum usage.

- Satellite's antenna patterns play an important role and must be designed to best cover the designated geographical area (which is generally irregular in shape).

- Satellites should be designed by keeping in mind its usability for short and long term effects throughout its life time.

- The earth station should be in a position to control the satellite if it drifts from its orbit ir is subjected to any kind of drag from the external forces.

- The designing of a satellite begins with a base line space craft design, meeting all the technical requirements such as EIRP and coverage. The synthesis process provides useful parameters such as size and weight of the spacecraft.

- Let us consider some major design considerations of a communication satellite.

7.2 DESIGN CONSIDERATIONS

7.2.1) Communication Considerations

- For telecommunication satellite, the main design considerations are:
  - i) Type of service to be provided
  - ii) Communication capacity
  - iii) Coverage area
  - iv) Technological limitations

- Depending upon the type of service to be provided be the satellite, basic specifications are laid down.
- For domestic fixed satellite services, the main parameters are EIRP per carrier, number of carriers and the assigned coverage area.
- For direct broadcast satellites, the number of television channels and coverage area is specified.
- Based on these parameters, satellites are designed to fulfill the areas needs and at the same time it should be made in the specified cost fulfilling all the technical constraints.
- While developing a satellite, the earth station's previous experience and in-house capabilities are also taken into account.
- Often, for same set of requirements, different types of configurations are often proposed.

7.2.2) Environmental Conditions

Different environmental conditions are encountered by a satellite during its mission. Some of them are mentioned below.

7.2.2.1) Zero Gravity:

- In geostationary earth orbit, effect of earth's gravity is negligible thus making the "zero gravity" effect.

- Disadvantage: This causes a problem for liquids to flow. The major issue of fuel is encountered. Thus an external provision has to be made to force the liquids to flow.
- Advantage: Absence of gravity leads to operation of deployment mechanism used for stowing antennas and solar panels during the launch.

7.2.2.2) Atmospheric pressure and temperature:

- At geostationary earth orbit, atmospheric pressure is very low, thus making the thermal conditions negligible which further leads to the increase in friction between surfaces.

- Thus additional lubricants are required to keep the satellite parts in motion.

- Due to the presence of electronic components inside the satellite, pressure us the satellite is higher making the functioning of the inner components of the satellite more manageable.

- Sun's heat also affects the external components of the satellite.

7.2.2.3) Space Particles:

- Besides planets, natural and artificial satellites, many other particles like cosmic rays, protons, electrons, meteoroids and manmade space debris exists in space.

- These particles collide with the satellites causing permanent damage to it and sometimes degrading the solar cells.

- Space debris, also known as orbital debris, space junk and space waste, is the collection of objects in orbit around Earth that were created by humans but no longer serve any useful purpose. These objects consist of everything from spent rocket stages and defunct satellites to explosion and collision fragments.

- The debris can include slag and dust from solid rocket motors, surface degradation products such as paint flakes, clusters of small needles, and objects released due to the impact of micrometeoroids or fairly small debris onto spacecraft. As the orbits of these objects often overlap the trajectories of spacecraft, debris is a potential collision risk.

- The vast majority of the estimated tens of millions of pieces of space debris are small particles, like paint flakes and solid rocket fuel slag. Impacts of these particles cause erosive damage, similar to sandblasting. The majority of this damage can be mitigated through the use of a technique originally developed to protect spacecraft from micrometeorites, by adding a thin layer of metal foil outside of the main spacecraft body.

- Impacts take place at such high velocities that the debris is vaporized when it collides with the foil, and the resulting plasma spreads out quickly enough that it does not cause serious damage to the inner wall. However, not all parts of a spacecraft may be protected in this manner, i.e. solar panels and optical devices (such as telescopes, or star trackers), and these components are subject to constant wear by debris and micrometeorites.

- The present means for spacecraft shielding, such as those used for the manned modules of the International Space Station, are only capable of protecting against debris with diameters below about 1 centimeter. The only remaining means of protection would be to maneuver the spacecraft in order to avoid a collision. This, however, requires that the orbit of the respective object be precisely known.

- If a collision with larger debris does occur, many of the resulting fragments from the damaged spacecraft will also be in the 1 kilogram mass range, and these objects become an additional collision risk.

- As the chance of collision is a function of the number of objects in space, there is a critical density where the creation of new debris occurs faster than the various natural forces that remove these objects from orbit. Beyond this point a runaway chain reaction can occur that quickly reduces all objects in orbit to debris in a period of years or months.

7.2.2.4) Magnetic Fields:

- Due to the magnetic field of earth, charged particles which are trapped in the surrounding region of the earth get deflected.

- This effect is more seen in the layers around the equator where the magnetic power of the earth is of maximum effect. This region is called the Van Allen's Belt.

- Even though satellites in geostationary earth orbit are not really affected by the earth's magnetic field, they have to pass through the Van Allen's belt during orbit raising (launching).

- The electric charges present in this belt affect the electronic components against radiation.

- To overcome this effect, large coils are used by satellites.

7.2.2.5) Other Considerations:

- During eclipses, satellite's solar cells face a loss of power and over the period of years, they degrade.

- Satellites in geostationary earth orbit are also affected by a number of external forces (like gravitational pull from sun and moon) which eventually makes them deviate gradually from their orbit.

- Moreover the solar radiation falling on the surface of the satellite generates pressure which gradually leads to a change in the eccentricity of the orbit.

- As satellite rotates full 360^0along its North – South axis over a sidereal day, the net effect of these forces on the satellite is to cause a gradual drift from its nominal position together with a short – term variations in pointing towards the earth. These perbutations are controlled by an on-board orbit control system.

7.3 LIFETIME AND RELIABILITY

7.3.1) Lifetime:

- The useful lifetime of a geostationary satellite is determined by the highest tolerable deviation in inclination and orbit location together with reliability of satellite's critical sub-system.

- A lifetime could be improved by increasing the fuel capacity and by saving fuel by accepting orbital deviation to the maximum extent that is possible. Saving fuel couldn't be implemented to a great level. So for this purpose propulsion is used.

- Propulsion: It is a method used to accelerate spacecraft and artificial satellites.

7.3.2) Reliability

- Reliability is counted by considering the proper working of satellites critical components. Reliability could be improved by making the critical components redundant. Components with a limited lifetime such as travelling wave tube amplifier etc should be made redundant.

- *Travelling Wave Tube Amplifier (TWTA):* travelling wave tube amplifiers have applications in both receiver and transmitter systems, and come in all shapes and sizes, but they all consist of three basic parts-the tube, the tube mount (which includes the beam focussing magnets) and the power supply.

- The main attraction of these devices is their very high gain (30-60 dB), linear characteristics and 1-2 octave bandwidth. They are quite widely used professionally, but are still rather scarce in amateur circles. This article describes a little of the theory of twts, and explains how to use them, in the hope that more amateurs may be able to acquire and use these fascinating components.

- When used as receiver RF amplifiers they are characterized by high gain, low noise figure and wide bandwidth, and are known as low noise amplifiers (LNAS). These usually come with tube, mount and power supply in one integral unit, with no external adjustments to make-just input socket, output socket and mains supply connections. A typical LNA has an octave bandwidth (eg 2-4 GHz), 30 dB gain, 8 dB noise figure, and a saturated power output of 10 mW, within a volume of 2 in by 2 in by 10 in.

- Transmitter TWTAs are naturally somewhat bulkier, and often have the poweror supplies as a separate unit. Medium-power tubes have outputs of up to about 10 W, while high-power tubes deliver several hundred watts. Such tubes have gains of the order of 30 or 40 dB, and bandwidths of up to an octave.

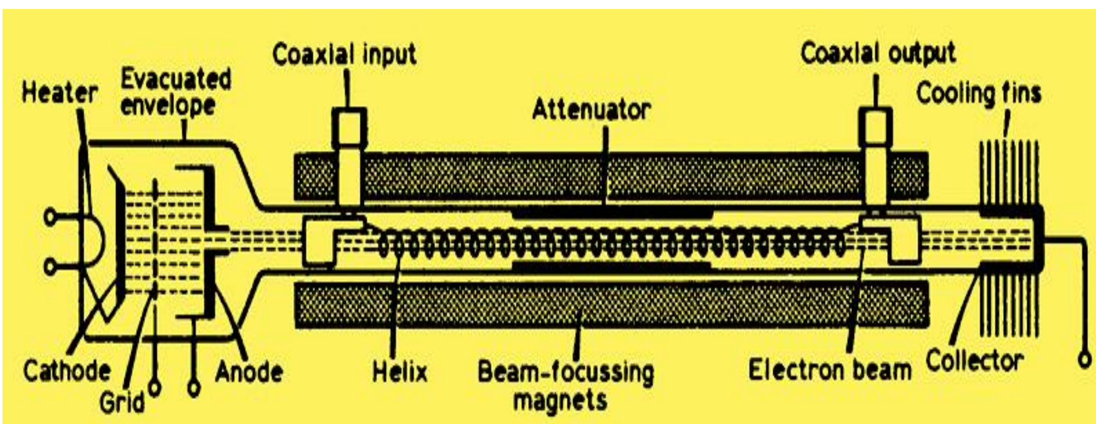


Figure 7.1: TWTA

- Other critical components are antenna reflectors, beaming assemblers etc.

- A reliability model is used to calculate the satellite's reliability. It is defined as "the probability that a given component or system performs its functions as desired within a specific time t.

- The failure rate for all components is calculated and they are categorized into the following three categories:

- Early high failure rate region: used for manufacturing faults, defects in material etc.
- Low failure: used for random component failure.
- High failure rate: used for components weave-out.

- Certainly early failures criteria is eliminated as most of the components are tested before used in the satellite.

- Random failures are more seen. They could be reduced by using reliable engineering techniques.

- The life-spam of component could be increased by improving manufacturing techniques and the type of material used to reduce the number of worn out parts and hence reducing the high failure rate criteria.

- It is sent that the failure rate is constant over time and is looking at this reliability can be determined.

- The system is made of several components, connected in a series, then the overall reliability is determined.

- By duplicating the less reliable and critical components, the overall reliability of the system could be improved. If any failure occurs in operational unit, then the standby unit takes over to develop a system with redundant components, its redundant elements are considered in parallel.

- Parallel redundancy is useful when the reliability of an individual sub-system is high.

- Example: consider a system having *i* parallel components in which reliability of each element is independent of others.

- If Qi is the unreliability of the $\mathbf{i^{th}}$ parallel element, then the probability that all units will fail is the product of the individual un-reliabilities:

$$Qs = Q1\ Q2\ Q3\ \ldots\ Qi$$

- When the un-reliability of all elements is equal, then Qs = Qi where Q is the un-reliability of each element.

- By doing a complete failure analysis, one could find out which failure occurs more than the rest and such analysis help in finding out the manufacturing defects in the product of a given batch of components or probably a design defect.

- This analysis is done to reduce the overall reliability to a value less than that predicted by the above analysis.

- Co-related failures could also be reduced by using units from different manufacturers. The design defects are generic to all satellite produced in a series. Generally these defects are detected and corrected to minimize their impact. This is done when a complete design change cannot be implemented.

- Even through the reliability can be improved by adding redundant devices and components, the weight of the satellite increases which again becomes a problem. Redundant component also increase the cost of the satellite.

- The two major cost components are:
    - Cost of equipment together with the switching and failure sensing mechanism used.
    - The associated increase in weight of the satellite resulting in an increased launch cost.

- Optimization techniques are performed for cost minimization purpose.

7.4 EXERCISE

1. Discuss the design considerations of a communication satellite.
2. What issues are faced by communication satellites wrt lifetime and reliability?
3. What are space particles? What is their impact on the satellites?
4. Explain the design and functionality of a travelling wave tube amplifier.

References

58. www.wikipedia.com
59. http://www.tech-faq.com/vsat.html
60. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
61. Rappaort, Wireless Communications Principals and Practices
62. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
63. P. Nicopolitidis ,Wireless Networks, John Wiley
64. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
65. http://www.britannica.com/EBchecked/topic/524891/satellite-communication

❖❖❖❖

8

COMMUNICATION SATELLITES – PART II

Contents:

8.1 SPACE CRAFT SUB-SYSTEMS

A communication satellite consists of two main functions, they are payload and bus. Payload is required for communication whereas bus is required for mechanical and electrical support. Bus supports altitude and orbit controls, propulsion, TT&C and electrical power where as payload supports the band used for communication, the space links and the devices to remove interferences.

8.2 PAYLOAD

The payload comprises of a Repeater and Antenna sub-system and performs the primary function of communication.

8.2.1) REPEATER

- It is a device that receives a signal and retransmits it to a higher level and/or higher power onto the other side of the obstruction so that the signal can cover longer distance.
- A repeater in the satellite receives the uplink RF signal and converts it to an appropriate downlink frequency.
- It does the work of processing the received signal. Two types of repeater architectures are used. They are given below.

8.2.1.1) Transparent Repeater

- It only translates the uplink frequency to an appropriate downlink frequency. It does so without processing the baseband signal. The main element of a typical transparent repeater is a single beam satellite. Signals from antenna and the feed system are fed into the low-noise amplifier through a bandpass filter.
- The bandpass filter attenuates all out-of band signals such as transmission from the ground stations of adjacent satellite systems. The low-noise amplifier provides amplification to the weak received signals.

- The spacecraft antenna is pointed towards a relatively warm earth having noise temperature about 300K. thus there is not much point in reducing the noise temperature below a certain point.

8.2.1.2) Regenerative Repeater

- A repeater, designed for digital transmission, in which digital signals are amplified, reshaped, retimed, and retransmitted. Regenerative Repeater can also be called as a device which regenerates incoming digital signals and then retransmits these signals on an outgoing circuit.

- It not only translates and amplifies the signal, but is also does the task of demodulation, baseband processing and demodulation. This architecture of repeater is the best suited for digital systems and it offers several advantages over transparent repeaters.

- When any digital signal is transmitted over a pair of wires, it degrades in amplitude. Regenerative repeaters receives the incoming signal, extracts the clock, then regenerates the original signal as a clean digital square wave as if it was the original signal transmitted from the source.

8.2.2) Antennas

- The function of an antenna of a space craft is to receive signals and transmit signals to the ground stations located within the coverage area of the satellite. The choice of the antenna system is therefore governed by the size and shape of the coverage area. Consequently, there is also a limit to the minimum size of the antenna footprint.

- An antenna (or aerial) is a transducer that transmits or receives electromagnetic waves. In other words, antennas convert electromagnetic radiation into electrical current, or vice versa. Antennas generally deal in the transmission and reception of radio waves, and are a necessary part of all radio equipment.

- Antennas are used in systems such as radio and television broadcasting, point-to-point radio communication, wireless LAN, cell phones, radar, andspacecraft communication. Antennas are most commonly employed in air or outer space, but can also be operated under water or even through soil and rock at certain frequencies for short distances.

- Physically, an antenna is an arrangement of one or more conductors, usually called *elements* in this context. In transmission, an alternating current is created in the elements by applying a voltage at the antenna terminals, causing the elements to radiate an electromagnetic field.

- In reception, the inverse occurs: an electromagnetic field from another source induces an alternating current in the elements and a corresponding voltage at the antenna's terminals. Some receiving antennas incorporate shaped reflective surfaces to collect the radio waves striking them and direct or focus them onto the actual conductive elements.

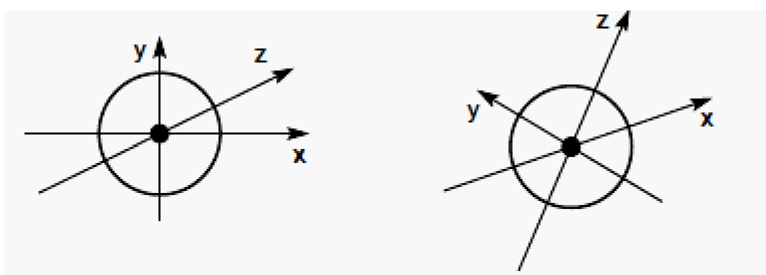
- 

Figure: 8.1: Radiation pattern of an isotropic radiator

- A theoretical reference antenna is the *isotropic radiator*, a point in space radiating equal power in all directions, i.e., all points with equal power are located on a sphere with the antenna as its center. The *radiation pattern*is symmetric in all directions (see figure 8.1).

  (Radiation Pattern: The radiation pattern of an antenna is the geometric pattern of the relative field strengths of the field emitted by the antenna. For the ideal isotropic antenna, this would be a sphere. For a typical dipole, this would be a toroid. The radiation pattern of an antenna is typically represented by a three dimensional graph, or polar plots of the horizontal and vertical cross sections.)

- This is a theoretical definition of an antenna, in real, the intensity radiation is not equal in all directions, and this is called *directive effects.* The simplest antenna is a *dipole antenna*, which can be made by a simple wire with a center-fed driven element. The current amplitude on such an antenna decreases uniformly from maximum at the center to zero at the ends.It's also called *Hertzian dipole* as this antenna was created by Heinrich Rudolph Hertz around1886.

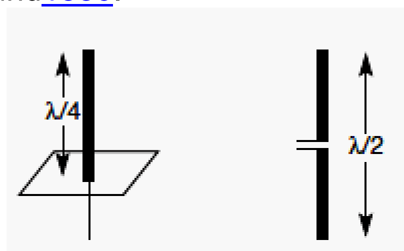


Figure 8.2: Simple Antenna

- The length of the dipole is not arbitrary, but, for example, half the wavelength *λ* of the signal to transmit results in a very efficient radiation of the energy. If mounted on the roof of a car, the length of *λ*/4 is efficient. This is also known as Marconi antenna. A *λ*/2 dipole has a uniform or *omni-directional*radiation pattern in one plane and a figure eight pattern in the other two planes as shown in Figure 8.3. This type of antenna can only overcome environmental challenges by boosting the power level of the signal.

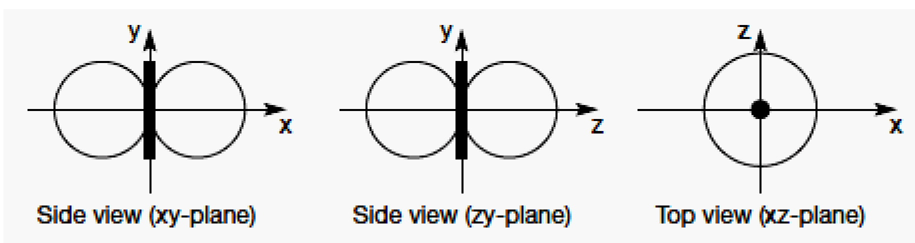


Figure 8.3: Radiation pattern of simple dipole

- Obstructions for these antennas could be mountains, valleys, buildings etc. as if they are placed between mountains or buildings, the omni-direction pattern doesn't work well. Thus, in these places, a *directional antenna* proves to be more effective. A *directional antenna* radiates greater power in one or more directions allowing for increased performance on transmit and receive and reduced interference from unwanted sources. Directional antennas provide increased performance over dipole antennas when a greater concentration of radiation in a certain direction is desired.
-

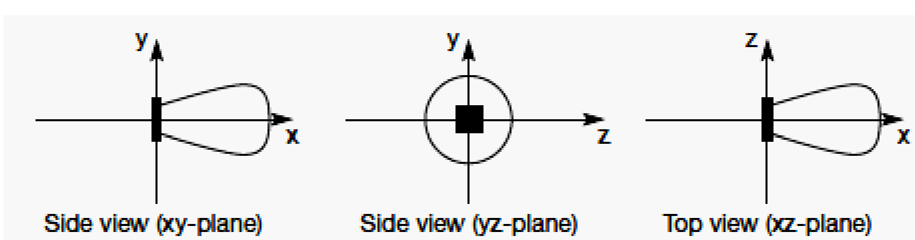


Figure 8.4: Radiation pattern of a directional antenna

- Many directional antennas can be combined together to form a *sectorized antenna*.

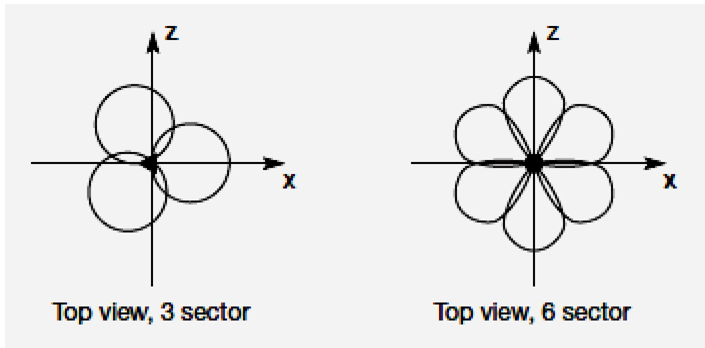


Figure 8.5: Radiation pattern of sectorized antenna

- Two or more antennas can also be combined to improve reception by counteracting the negative effects of multi-path propagation. These antennas, also called *multi-element antenna arrays*, allow different diversity schemes. One such scheme is *switched diversity*or *selection diversity*, where the receiver always uses the antenna element with the largest output.

- *Diversity combining*constitutes a combination of the power of all signals to produce gain. The phase is first corrected to avoid cancellation. As shown in Figure 8.6, different schemes are possible. On the left, two $\lambda/4$ antennas are combined with a distance of $\lambda/2$ between them on top of a ground plane. On the right, three standard $\lambda/2$ dipoles are combined with a distance of $\lambda/2$ between them. Spacing could also be in multiples of $\lambda/2$.

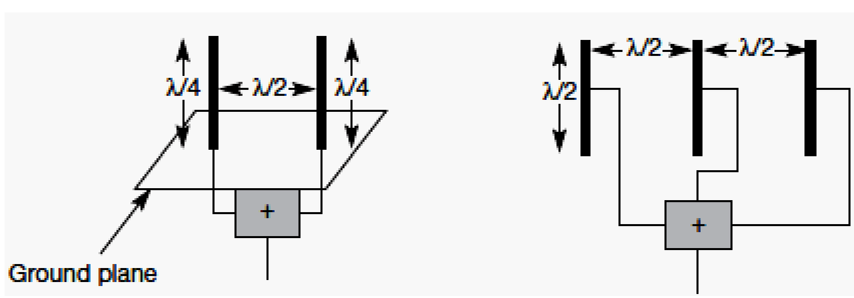


Figure 8.6: Diversity antenna systems

- Another type of antennas is *smart antennas*, which combine multiple antenna elements with signal processing to optimize the radiation/reception pattern in response to the signal environment. These antennas can adapt to changes in reception power, transmission conditions and many signal propagation effects. They are also known as also known as adaptive array antennas. They use smart signal processing algorithms used to identify spatial signal signatureto track and locate the antenna beam on the mobile/target.

- It would not just be base stations that could follow users with an individual beam. Wireless devices, too, could direct their electromagnetic radiation, e.g., away from the human body towards a base station. This would help in reducing the absorbed radiation.

(Antenna gain relates the intensity of an antenna in a given direction to the intensity that would be produced by a hypothetical ideal antenna that radiates equally in all directions and has no losses. Since the radiation intensity from a lossless isotropic antenna equals the power into the antenna divided by a solid angle of 4π, we can write the following equation:

$$Gain = 4\pi \left(\frac{\text{Radiation Intensity}}{\text{Antenna Input Power}} \right)$$

Although the gain of an antenna is directly related to its directivity, the antenna gain is a measure that takes into account the efficiency of the antenna as well as its directional capabilities. In contrast, directivity is defined as a measure that takes into account only the directional properties of the antenna and therefore it is only influenced by the antenna pattern. However, if we assumed an ideal antenna without losses then Antenna Gain will equal directivity as the antenna efficiency factor equals 1 i.e. 100% efficiency. In practice, the gain of an antenna is always less than its directivity.**)**

8.3 BUS

The bus or payload platform consists of the subsystems that support the payload. These subsystems typically include:

- ***Structures subsystem*****:** the physical structure of the spacecraft, to which all electronics boxes, thrusters, sensors, propellant tanks, and other components are mounted;

- ***Electric power/distribution subsystem (EPS or EPDS):*** the hard- and software used to generate and distribute electrical power to the spacecraft, including solar arrays, batteries, solar-array controllers, power converters, electrical harnesses, battery-charge-control electronics, and other components;

- ***Telemetry, tracking, and command subsystem (TT&C):*** The electronics used to track, monitor, and communicate with the spacecraft from the ground. TT&C equipment generally includes receivers, transmitters, antennas, tape recorders, and state-of-health sensors for parameters such as temperature, electrical current, voltage, propellant tank pressure, enable/disable status for various components, etc.;

- ***Propulsion subsystem:*** Liquid and solid rockets or compressed-gas jets and associated hardware used for changing satellite attitude, velocity, or spin rate. Solid rockets are usually used for placing a satellite in its final orbit after separation from the launch vehicle. The liquid engines (along with associated plumbing lines, valves, and tanks) may be used for attitude control and orbit adjustments as well as final orbit insertion after launch;

- ***Power supply:*** The primary electrical power for operating electronic equipment is obtained from solar cells. Individual cells can generate small amounts of power, and therefore array of cells in series-parallel connection are required. Cylindrical solar arrays are used with spinning satellites, thus the array are only partially in sunshine at any given time. Another type of solar panel is the rectangular array or solar sail. solar sail must be folded during the launch phase and extended when in geo-stationary orbit. Since the full component of solar cells are exposed to sun light ,and since the Sail rotate to track, the sun , they capable of greater power output than cylindrical arrays having a comparable number of cells.To maintain service during an eclipse, storage batteries must be provided.

- ***Attitude control:*** The attitude of a satellite refers to its Orientation in space. Much of equipment carried abroad a satellite is there for the purpose of controlling its attitude. Attitude control is necessary, for example, to ensure that directional antennas point in the proper directions. In the case of earth environmental satellites the earth-sensing instrument must cover the required regions of the earth, which also requires attitude control. A number of forces, referred to as disturbance forces can alter attitude, some examples being the gravitational forces of earth and moon, solar radiation, and meteorite impacts.

- ***Station keeping:*** A satellite that is normally in geo-stationary will also drift in latitude, the main perturbing forces being the gravitational pull of the sun and the moon. The force causes the inclination to change at the rate of about 0.85 deg/year. If left uncorrected, the drift would result in a cycle change in the inclination going 0 to 14.67deg in 26.6 years and back to zero, when the cycle is repeated. To prevent the shift in inclination from exceeding specified limits, jets may be pulled at the appropriate time to return the inclination to zero. Counteracting jets must be pulsed when the inclination is at zero to halt that change in inclination.

- ***Thermal control:*** Satellites are subject to large thermal gradients, receiving the sun radiation on one side while the other side faces into space. In addition, thermal radiation from the earth, and the earth's abedo, which is the fraction on the radiation falling on the earth which is reflected can be sight for low altitude, earth-orbiting satellites, although it is negligible for geo-stationary satellites. Equipment in the satellite also generates heat which has to be removed. The most important consideration is that the satellite's equipment should operate as near as possible in a stable temperature environment. Various steps are taken to achieve this. Thermal blankets and shields may be used to provide insulation. Radiation mirrors are often used to remove heat from communication payload. These mirrored drums surrounded the communication equipment shelves in each case and provide good radiation paths for the generated heat to escape in to surround space.To maintain constant-temperature conditions, heaters may be switched on to make up for the hearts may be switched on to make reduction that occurs when transponders are switched off.

8.4 TELEMETRIC TRACKING AND COMMAND SUBSYSTEM

The main functions of TT&C are:
1) Monitor the performance of all the satellite sub-systems and transmit the monitored data to the satellite control center.
2) Support the determination of orbital parameters.
3) Provide a source earth station for tracking.
4) Receive commands from the control center for performing various functions of the satellite.

Telemetry system

- The telemetry, tracking, and command (TT&C) subsystem performs several routine functions abroad a spacecraft. The telemetry or "telemetering" function could be interpreted as "measurement at a distance". Specifically, it refers to the overall operation of generating an electrical signal proportional to the quantity being measured, and encoding and transmitting this to a distant station, which for satellite is one of the earth stations, which for the satellite is one of the earth stations.

- Data that are transmitted as telemetry signals include attribute information such as obtained from sun earth sensors; environmental information such as magnetic field intensity and direction; the frequency of meteorite impact and so on ;and spacecraft information such as temperatures and power supply voltages, and stored fuel pressure.

- Summary of the parameters monitored by the Telemetry system are:
  1) Voltage, current and temperature of all major sub-systems.
  2) Switch status of communication transponders.
  3) Pressure of the propulsion tanks
  4) Outputs from altitude sensors.
  5) Reaction wheel speed

Command systems

- Command system receives instructions from ground system of satellite and decodes the instruction and sends commends to other systems as per the instruction.

- Example of commands are:
  1) Transponder switching
  2) Switch matrix configuration
  3) Antenna pointing control
  4) Controlling direction and speed of solar array drive
  5) Battery reconditioning
  6) Beacon switching
  7) Thruster firing
  8) Switching heaters of the various sub-systems

Tracking

- Tracking of the satellite is accomplished by having the satellite is accomplished by having the satellite transmit beacon signals which are received at the TT&C earth stations. Tracking is obviously important during the transmitter and drift orbital phases of the satellite launch.

- When on-station, a geo-stationary satellite will tend to shifted as a result of the various distributing forces, as described previously. Therefore it is necessary to be able to track the satellites movements and send correction signals as required. Satellite range is also required for time to time. This can be determined by measurement of propagation delay of signals specially transmitted for ranging purposes.

8.5 EXERCISE

1. Discuss the TT&C system of a communication satellite.
2. Elaborate on the bus system of the communication satellites.
3. Write a note on antennas used for communication satellites.

4. What is a Regenerative Repeater? How is it different from Transparent Repeater?

References

66. www.wikipedia.com
67. http://www.tech-faq.com/vsat.html
68. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
69. Rappaort, Wireless Communications Principals and Practices
70. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
71. P. Nicopolitidis ,Wireless Networks, John Wiley
72. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
73. http://www.britannica.com/EBchecked/topic/524891/satellite-communication

❖❖❖❖

9

MULTIPLE ACCESS TECHNIQUES

Contents:

9.1 INTRODUCTION

- Multiple accesses is defined as the technique where in more than one pair of earth stations can simultaneously use a satellite transponder.It is a technique used to explore the satellite's geometric advantages and is at the core of satellite networking.

- In a wireless communication system, radio resources must be provided in each cell to assure the interchange of data between the mobile terminal and the base station. Uplink is from the mobile users to the base station and downlink is from the base station to the mobile users. Each transmitting terminal employs different resources of the cell. A multiple access scheme is a method used to distinguish among different simultaneous transmissions in a cell. A radio resource can be a different time interval, a frequency interval or a code with a suitable power level.

- All these characteristics (i.e., time, frequency, code and power) univocally contribute to identify a radio resource. If the different transmissions are differentiated only for the frequency band, we have the Frequency Division Multiple Access (FDMA). Whereas, if transmissions are distinguished on the basis of time, then they are consider the Time Division Multiple Access (TDMA). Finally, if a different code is adopted to separate simultaneous transmissions, we have the Code Division Multiple Access (CDMA).

- However, resources can be also differentiated by more than one of the above aspects. Hence, hybrid multiple access schemes are possible (e.g., FDMA/TDMA). In a satellite communication, radio resources can be re-used between sufficiently far stations, provided that the mutual interference level is at an acceptable level.

- This technique is adopted by FDMA and TDMA air interface, where the reuse is basically of carriers. In the CDMA case, the number of available codes is so high that the code reuse among cells (if adopted) does not increase the interference. In uplink, a suitable Medium Access Control (MAC) protocol is used to regulate the access of different terminals to the resources of a cell that are provided by a multiple access scheme.

In this unit, we concentrate on all the three schemes of multiple access.

9.2 FREQUENCY DIVISION MULTIPLE ACCESS (FDMA)

- FDMA is a channel access method used in multiple-access protocols as a channelization protocol. FDMA gives users an individual allocation of one or severalfrequency bands, or channels. Multiple Access systems coordinate access between multiple users.

Key features of FDMA are:

- FDMA requires high-performing filters in the radio hardware, in contrast to TDMA and CDMA.
- FDMA is not vulnerable to the timing problems that TDMA has. Since a predetermined frequency band is available for the entire period of communication, stream data (a continuous flow of data that may not be packetized) can easily be used with FDMA.
- Due to the frequency filtering, FDMA is not sensitive to near-far problem which is pronounced for CDMA.
- Each user transmits and receives at different frequencies as each user gets a unique frequency slot

- It is important to distinguish between FDMA and frequency-division duplexing (FDD). While FDMA allows multiple users simultaneous access to a certain system, FDD refers to how the radio channel is shared between the uplink and downlink (for instance, the traffic going back and forth between a mobile-phone and a base-station).

- Furthermore, frequency-division multiplexing (FDM) should not be confused with FDMA. The former is a physical layer technique that combines and transmits low-bandwidth channels through a high-bandwidth channel. FDMA, on the other hand, is an access method in the data link layer.

- FDMA also supports demand assignment in addition to fixed assignment. Demand assignment allows all users apparently continuous access of the radio spectrum by assigning carrier frequencies on a temporary basis using a statistical assignment process. The first FDMA demand-assignment system for satellite was developed by COMSAT for use on the Intelsat series IVA and V satellites.

- In this scheme, a bandwidth is assigned to an earth station and is divided into n segments to manage the network traffic.

  FDMS is divided into two categories:
  - Multiple Channel Per Carrier
  - Single Channel Per Carrier

A) Multiple Channel Per Carrier (MCPC)

- Each base band filter is an earth station receiver, which correspondent to a specific transmitting station.
- Changes in the traffic are not favorable.

- MCPC is referred as Frequency Division Multiplexing / Frequency Modulation/ Frequency Division Multiple Access scheme.

B) Single Channel Per Carrier

- Certain applications have a low requirement. At such times, the earth station's load is comparatively less. Here, instead of assigning multiple channel carriers to each station demand assign or pre-assigned schemes are preferred and used.

- In demand-assign, pool of frequencies is shared by each earth station which they use as and when required by requesting a channel from the pool manager.

- In pre-assigned method, channels are permanently assigned to an earth station for its use.

- Demand-assign is more preferred over pre-assigned method, as a reduction in cost is possible through sharing of equipments.

- The frequency pool is managed in two ways: Distributed and Centralized.

- *Distributed:* Each earth station can obtain a channel from the pool on its own i.e. without taking the permission from the pool manager. This scheme is more reliable as failure of one earth station will not affect the entire network. Nevertheless, this scheme is less cost efficient due to increase complexity in earth station's working in the network.

- *Centralized:* this scheme is simple to control and it provides a higher usage of channels because of the availability of all information at a single point. This scheme also offers a lower time connection time. Nevertheless, this network is less reliable, as it is prone to a single point failure. If the pool manager crashes, the entire system fails.

- It is also possible to have a hybrid frequency management schemes in which the network provides acombination of Distributed and Centralized frequency management functions to get the advantage of both the systems.

9.2.1) Design Considerations of FDMA system

Design considerations of FDMA system are as follows:

A) Impairments caused by Satellite's High Power Amplifiers:

- High power amplifiers in satellites face the problem of non-linearity when the output level approaches saturation. This leads to inter modulation noise.

  (Inter modulation noise: It is the unwanted amplitude modulation of signal containing two or more different frequencies).

- To minimize the effect of inter modulation noise, the derive level of the output stage of satellite transmitter is made to *back off* (i.e. its power is reduced). Thus the full power capacity of the amplifier is not utilized causing reduction in the capacity.

B) Other Impairments:

- To increase the frequency utilization, carriers are brought as close to each other as possible. This causes the spectrum to overlap, leading to *adjacent channel interference.*
  (Spectrum: It is an ordered array of the components of an emission / wave)

- Convolution noise can occur when adjacent carrier has smaller amplitude than the desired carrier. Impulse noise can occur when adjacent carrier has greater amplitude than the desired carrier.

- The above interferences can be minimized by filtering the out of the band inter modulations in each transponders.

  (Transponder: It is an electrical device designed to receive a specific signal and it automatically transmits a specific reply.)

- Delay occurs in filters in transmission path leading to another source of distortion. This again increases the noise.

9.2.2) Transponder Utilization

- Two factors limit the number of FDMA accesses through a transponder.
  i) Inter modulation noise
  ii) Spectrum utilization efficiency

9.2.3) Summary of salient features of FDMA

A) Advantages of FDMA:

- Uses existing hardware and hence this technology is cost efficient
- Network timing is not required, hence making the system less complex.
- No restrictions regarding the type of baseband type of modulation is there.

B) Disadvantage of FDMA

- Inter modulation noise in the transponder leads to interference with other links sharing new spectrum and thus reduces the capacity of satellite.
- Flexibility in channel allocation is less (as seen in MCPC, but not in SCPC).
- Uplink power control is required to maintain the link quality.
- As strong and weak carriers, both are used, weak carriers are often suppressed.

9.3 TIME DIVISION MULTIPLE ACCESS (TDMA)

- Time division multiple access (TDMA) is a channel access method for shared medium networks. It allows several users to share the same frequency channel by dividing the signal into different time slots. The users transmit in rapid succession, one after the other, each using his own time slot. This allows multiple stations to share the same transmission medium (e.g. radio frequency channel) while using only a part of its channel capacity.

- TDMA is a type of Time-division multiplexing, with the special point that instead of having one transmitter connected to one receiver, there are multiple transmitters. In the case of the uplink from a mobile phone to a base station this becomes particularly difficult because the mobile phone can move around and vary the timing advance required to make its transmission match the gap in transmission from its peers.

Features of TDMA

- Shares single carrier frequency with multiple users
- Non-continuous transmission makes handoff simpler
- Slots can be assigned on demand in dynamic TDMA
- Less stringent power control than CDMA due to reduced intra cell interference
- Higher synchronization overhead than CDMA
- Advanced equalization may be necessary for high data rates if the channel is "frequency selective" and creates inter-symbol interference
- Cell breathing (borrowing resources from adjacent cells) is more complicated than in CDMA
- Frequency/slot allocation complexity
- Pulsating power envelop: Interference with other devices

For satellite communication TDMA works in the following manner.

- TDMA systems are used in commercial satellite applications. The first system type is the classic TDMA implementation employing a single modulated carrier occupying the full transponder's bandwidth.

- This system is most common for TDMA networks and is also most efficient from a capacity standard's point.
- Each user is allocated a specific time slot for transmission due to which overlapping is avoided.
- System capacity is increased as only a single carrier is present at any given time.
- Disadvantage is that the messages need to be stored, compressed and transmitted during one or more specific time slots. At network level, all transmissions must be synchronized to avoid collision between the bursts.

(Burst: It is a term used in a number of information technology contexts to mean a specific amount of data sent or received in one intermittent time.)

- An earth station has a full access to a transponder during its allocated time slot.
- Transmissions are in frame format.
- On receiving al the bursts, earth station removes the data addresses from it.
- Guard time is used to separate time-slots. The time-slot size depends on the traffic requirements.
- TDMA also works on demand- assign method.
- Reference burst has three parts

Carrier and Bit Recovery	**Unique Word**	**Control Bits**
Used by the receivers for recovering the carrier and bit time which is essential for coherent demodulation	It is used for burst synchronization. This is done by co-relating a stored replica of the unique word with the received bits.	Control bits contain information such as station ID and engineering service messages for network management.

- Carrier and bit recovery, unique word and control bits together are called the *preamble*.
- CDMA uses two methods for synchronization:

9.3.1) Design Considerations

Design considerations of TDMA system are as follows:

I) Closed Loop:

- Whatever adjustments are made to the burst's position is based on the real-time measurements.

  Transmission begins at *i*
  Assigned time slot is acquired.
  Estimated position of the initial burst is *Tm*
  Transmission is done at lower power.

Received burst comparison with the desired position is done- *Ta*
Time-slot correction *(Ta – Tm)* is done.

- A carried modulated with a pseudo random signal is transmitted. Such a signal has a noise like property and when transmitted at low power does not affect the other bursts.

- Receiver has a co-relation receiver where a correction is equal to the difference between the received and desired burst position is used. This method is applied to the original burst position to obtain the current burst position.

II) ***Frame Efficiency:*** It is defined as the ratio of time devoted for useful transmission to the total frame length in a TDMA system.

III) Transponder Utilization:

- Transponder utilization depends upon the EIRP (Equivalent isotropically radiated power) or the bandwidth.

- Maximum permissible bit rate is governed by the available transponder bandwidth possible when EIRP is sufficient.

9.3.2) Summary of salient features of TDMA

I) Advantages:

- Here, satellite power utilizations can be maximized as inter modulation noise is minimum.
- Uplink power control is not required.
- Transmission plans and capacity management is done by the satellite are very flexible.
- The digital format of TDMA allows utilization of all advantages of digital techniques.

II) Disadvantages

- It requires a network wide time synchronization which makes the entire system very complex.
- Analog of digital conversions are required.
- Interface with analog terrestrial plan is expected.

9.4 FDMA & TDMA

- In situations where connectivity is required between multiple spot beams, then routing of signal to an appropriate beam is done by having frequency-to-beam correspondence.

- Sub-bands are made and each of them provides a unique route between two spot beams.

- Here transponders can be accessed in FDMA or TDMA mode. As each earth station have to hop between transponders to route the traffic to the desired spot beam this technique is called transponder hopping.
- For *n* spot beams, n^2transponders are required.
- To make this work, flexibility in altering the frequency bands is required. This is done by using switchable routing method.
- In switchable routing method, channels are switched as desired in order to change the available bandwidth of each beam.
- A programmable switch located on the satellite router bursts to spot beams according to a set plan.
- The earth station can direct its transmission to any spot by transmitting in the appropriate time slot.
- Beams are arranged in non-overlapping time-slots.

9.5 CODE DIVISION MULTIPLE ACCESS (CDMA)

- CDMA uses a modulation technique called spread spectrum.Here all the users transmit signals simultaneously on the multiple access schemes.(Spread Spectrum: It refers to a modulation technique that converts the baseband signal to a modulated signal with a spectrum bandwidth that covers or is spread over the band orders of magnitude larger than that normally necessary to transmit the baseband signal itself.)
- It could be used as a multiple access system by giving each user a unique pseudo random code rather than a unique carrier frequency or time slot.
- All the users contribute to the noise background.
- To detect the desired signal in the presence of all the interferences, the composite signal is cross-correlated with the known pseudo random number spreading sequence.
- The net performance is improved essentially by the ration of the un-spread signal bandwidth.

9.5.1) Features of CDMA are:

- Highly resistant to interferences and thus satellite spacing could be reduced considerably without causing unacceptable degradation in the received signal quality.

- Spread spectrum sequences are resistant to multiple noises present in the mobile terminals.

- Small antennas can be used without any interference issues from the neighboring satellites.

- CDMA is a very secure form of communication.

9.5.2) Implementing CDMA

- CDMA technique could be implemented in two forms: Direct sequence spread spectrum and Frequency hopping spread spectrum.

- The property of pseudo random sequences is used by CDMA.

I) Pseudo-Random Sequences

- They are a set of signals which appear ti be a set of random sequences. They are repeated over a time interval, say of Tr.

- In order to use such pseudo (false) random sequences signals in digital form, shift registers are required.These shift registers could be used for maximum length of code of the value $p=(2^{m-1})$,where m-bit register is used.

- This is also called maximum length linear shift register sequence.

- These sequences have 2^{m-1} *ones* and 2^{m-1}*-1zeros* that are places randomly, thus making the entire sequence look random.

II) Direct Sequence Spread Spectrum (DSSS)

- Definition: Direct sequence spread spectrum is a modulation technique where the transmitted signal takes up more bandwidth then the information signal that is being modulated.

- Direct sequence spread spectrum transmissions multiply the data being transmitted by a ‘noise’ signal.

- The noise signal is the pseudo random sequence and has a frequency much higher than that of the original signal. It thus spreads the energy of the original signal into a much wider band

- The resultant signal appears like noise which could be reconstructed to the original signal at the receiving end by multiplying it by the same pseudo random sequence.This process is known as de-spreading. For de-spreading to work correctly, the transmitter and receiver must be synchronized.

- Sometimes while sending the signal from the transmitter's end, other noises like inter modulation noise and thermal noise are transmitted to the receiver. This is also called as narrow-band interference.

- The receiving signal is given as:

  $R_x (t) = C_1 (t) + C_2 (t) + \ldots + C_n (t)$

  Where: $C_n (t)$ is the received signal from the n^{th} transmitter n (t) is the system noise

III) Frequency Hopping Spread Spectrum (FHSS)

- It is a method of transmitting radio signals by rapidly switching a carrier among many frequency channels, using the pseudo random sequences (which are known to both transmitter and receiver).

- Frequency hopping spread spectrum offers three main advantages over fixed frequency transmission techniques:
  i) Spread spectrum signals are highly resistant to narrow band interferences. The process of re-collecting a spread signal spreads out the interfering signal, causing t to recede into the background.

  ii) Spread spectrum signals are difficult to intercept. Frequency hopping spread spectrum signals simply appears as an increase in the background noise to a narrow band receiver.

  iii) Spread spectrum transmission can share a frequency band with many types of conventional transmissions with minimal interference.

- Interference in frequency hopping spread spectrum is caused at instants when an unwanted signal appears within the pass band of the desired signal. It can occur under following conditions:

  i) Transmission of other users of multiple-access channel falls within the range of receiver's pass band.

  ii) Inter modulation noise can be generated due to non-linearities of receiver's channels.

- Interference is noise like when hopping rate is much higher than the information rate. Interference is coherent when hopping rate is smaller than the information rate.

9.6 ACCESS PROTOCOLS FOR DATA TRAFFIC

- All the previous schemes were dealing only with continuous streams of data. The other forms of transfers include the transfer of large amounts of data, channel allocation in demand-assign systems etc.

- Such transfers are categorized as “burst of high activity”. Here the system should be prepared for larger data traffic and delays in message deliveries. Hence, to overcome this problem, allocation of capacity of *pre message* or *packet basis* should be used.

- Accessing schemes for data traffic are categorized as following:
    - Channel reservation scheme
    - Contention protocols

- Packet reservation protocols

9.6.1) Channel Reservation Scheme

- Here the channels are reserved according to the duration of each message transmission. Channels can be pre-assigned or demand-assigned for a particular transaction.

- TDMA and FDMA schemes use this kind of technique.

- When the traffic load is high on the network, then the channel connection time becomes comparable to message transfer time, otherwise its negligible.

- The upper and the lower bound the channel throughput is calculated and it is observed for fixed assigned schemes, where the throughput reaches the maximum value, then the traffic becomes less busty and vice versa. (Throughput is the average rate of successful message delivery over a common channel).

- Generally the inter-arrival message delay does not effects the throughput of the demand-assigned scheme but it is affected by the delay factors on the message.

9.6.2) Contention Protocol

- Here, each use accesses multiple access channels without co-coordinating with any other user that is using the network. Due to this, collisions occur, leading to the loss of packets (data).

- Thus, a re-transmission of lost packets is done and again the network channel is utilized.

- To reduce the probability of collision, various schemes are applied. Such schemes which are suitable for large population of busty users with very short messages are discussed below.

9.6.3) ALOHA Scheme

- This scheme is also known as random access protocol as no coordination is required between the users. Terminals can transmit their data regardless of the activity of other terminals. If the message is successful, then the base station sends an acknowledgement (over the feedback channel).

- If the terminal does not receive an acknowledgement, it retransmits the message after waiting for a random amount of time.

- Here the delay is mainly determined by the probability that a packet is not received (which is because of the interference from other transmissions leading to collision) and the average value of random waiting time before transmission is considered.

9.6.4) Reservation ALOHA

- The main difference between slotted ALOHA and reservation ALOHA is that, any slot is available for utilization without regard to its prior usage in the slotted ALOHA scheme.

- In Reservation ALOHA contention based reservation scheme.A slot is temporarily considered to be 'owned' by a station that successfully used it. Once the station has completed its transmission, it simply stops sending the data. Here the idle slots are considered to be available to all the stations which may then implicitly reserve the slot on contention basis.

- This scheme has shorter delays and it efficiently supports higher level of utilization.

9.6.5) Slot Reservation ALOHA

- This extension of the slotted-ALOHA scheme allows time-slots to be reserved for transmission by an Earth station. This can be achieved implicitly, by which the transmitting station initially contends for an available slot with other transmitting terminals.

- Available slot locations are made known to all transmitting stations within the network by the network control station using a broadcast channel. Once a transmitting station successfully gains access to a particular slot by contention, the network controller informs all other transmitting stations that the slot is no longer available, and the successful transmitting station retains the slot until transmission is complete.

- The network controller then informs all stations on the network that the slot is available for contention once more. The other means of slot reservation is achieved explicitly, whereby a transmitting station requests the network to reserve a particular slot prior to transmission. In general terms, this mode of operation is termed a packet reserved multiple access scheme (PRMA).

9.7 EXERCISE

1. Discuss channel reservation scheme for Satellite communication.
2. Explain the principle used in spectrum spreading and dispreading. How is this used to minimize interference in a CDMA system?
3. Explain the technique of TDMA. How TDMA network is advantageous over FDMA network?
4. Discuss the features of CDMA.
5. Differentiate DSSS with FHSS.

References

74. www.wikipedia.com
75. http://www.tech-faq.com/vsat.html
76. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
77. Rappaort, Wireless Communications Principals and Practices
78. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
79. P. Nicopolitidis ,Wireless Networks, John Wiley
80. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
81. http://www.britannica.com/EBchecked/topic/524891/satellite-communication.

❖❖❖❖

10

EARTH STATIONS – PART I

Contents:

10.1 INTRODUCTION

- Earth Stations are a vital element in any satellite communication network. The function of an earth station from and transmit information to the satellite network in the most cost efficient and reliable manner while retaining the desired signal quality.

- Depending on the earth application, an earth station may have both transmit and receive capabilities or may only have both transmit and receive capabilities or may only be capable of either transmission or reception.

- Further categorization can be based on the kind of service provided. The fundamental parameter is describing an earth station in the carrier to noise ratio.

10.2 DESIGN CONSIDERATIONS

The design considerations depend on a number of factors, some of them are:

1) Type of service: fixed satellite service, mobile satellite service or broadcast satellite service.

2) Type of communication requirements: telephony, data, television etc.

3) Required baseband signal quality at the destination.
4) Traffic requirements: number of channels, type of traffic-continuous or bursty
5) Cost
6) Reliability

10.2.1) International Regulations

- Most of the fixed satellite service frequency bands are shared with the terrestrial systems. For them to coexist, the International Telecommunication Union (ITU) has specified certain constraints in the transmitted effective radiated power (EIRP) of satellites.

- By limiting the EIRP of satellite for applications like direct broadcast and mobile communication, a smaller diameter antenna could be used. This leads to exclusive allocation of frequency bands. The limitations seen in these applications are mainly because of the technological constraints of the space and ground segments.

10.2.2) Technical Constraints

- The transmitter power of a satellite is limited by the maximum DC power which a satellite can generate and the upper limit of the reliable power amplifiers. The maximal spacecraft antenna gain is limited by the practical constraints imposed on the satellite antenna diameter, if the gain falls for a given antenna size with the decrease in the frequency and therefore the EIRP limitation is more acute at lower frequencies.

- Technical constraints apply to the earth stations hardware and software. Factors which need to be included are the cost for *antennas at earth station, QoS to be maintained, cost of other equipments, floor area, environmental factors, interference considerations and recurring costs.*

10.3) INTRODUCTION TO GENERAL CONFIGURATIONS

Any earth station consists of four major subsystems: Transmitter, Receiver, Antenna, Tracking equipment
Two other important subsystems are

1) Terrestrial interface equipment
2) Power supply.

The earth station depends on the following parameters
i) Transmitter power
ii) Choice of frequency
iii) Gain of antenna
iv) Antenna efficiency
v) Antenna pointing accuracy

vi) Noise temperature
vii) Local conditions such as wind, weather etc,
viii) Polarization
ix) Propagation losses

The functional elements of a basic digital earth station are shown in the below figure 10.1

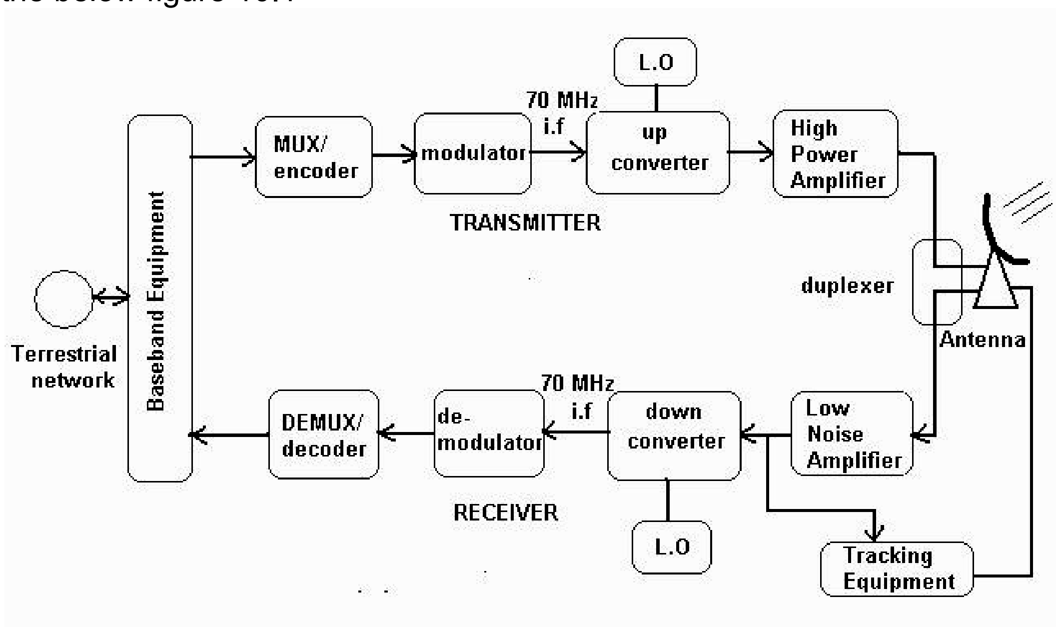


Figure 10.1 A general configuration of an earth station

- Digital information in the form of binary digits from terrestrial networks enters earth station and is then processed (filtered, multiplexed, formatted etc.) by the base band equipment.

- The encoder performs error correction coding to reduce the error rate, by introducing extra digits into digital stream generated by the base band equipment. The extra digits carry information. The presence of noise and non-ideal nature of any communication channel produces error rate is established above which the received information is not stable.

- The function of the modulator is to accept the symbol stream from the encoder and use it to modulate an intermediate frequency (I.F) carrier. In satellite communication, I.F carrier frequency is chosen at 70 MHz for communication using a 36 MHz transponder bandwidth and at 140 MHz for a transponder bandwidth of 54 or 72 MHz. The I.F is needed because it is difficult to design a modulator that works at the uplink frequency of 6 GHz (or 14GHz) directly.

- The modulated I.F carrier is fed to the up-converter and frequency-translated to the uplink r-f frequency.

- This modulated R.F carrier is then amplified by the high power amplifier (HPA) to a suitable level for transmission and radiation by the antenna to the satellite.

- On the receive side, the earth station antenna receives the low-level modulated R.F carrier in the downlink frequency spectrum.

- The low noise amplifier (LNA) is used to amplify the weak received signals and improve the signal to Noise ratio (SNR). The error rate requirements can be met more easily.

- R.F is to be reconverted to I.F at 70 or 140 MHz because it is easier design a demodulation to work at these frequencies than 4 or 12 GHz. The demodulator estimate which of the possible symbols was transmitted based on observation of the received if carrier.

- The decoder performs a function opposite that of the encoder. Because the sequence of symbols recovered by the demodulator may contain errors, the decoder must use the uniqueness of the redundant digits introduced by the encoder to correct the errors and recover information-bearing digits.

- The information stream is fed to the base-band equipment for processing for delivery to the terrestrial network.The tracking equipments track the satellite and align the beam towards it to facilitate communication.

10.3.1)Antenna Systems

Most of the earth stations use reflector antennas as these antennas provide high gain and desirable side lobe characteristics. The antenna system options are

1. Large antenna: say, for INTELSAT earth station typical diameter: 30M(Cassegrain geometry used)

2. Small antenna: say, for option o0f direct broad television (DBS – TV). For deep space communication, the diameter of antenna may be very large, say over 35m.

The efficient utilization of two natural resources- the radio spectrum and the geostationary orbit- are affected by the side lobe characteristic.

10.3.2) Asymmetric Configurations

In an axi-symmetric configuration the antenna axes are symmetric with respect to the reflector, which results in a relatively simple mechanism structure and antenna mount. The axi-symmetric antenna configuration has been used very widely until recently. Depending on the feed arrangement, several types of configurations are possible. Two most commonly used arrangements are:

i. Prime Focus Feed:

- It consists of a parabolic reflector antenna which is fed from a primary feed source located at the focus of the parabolic reflector. Owing to the geometry of the arrangement, the signal reflected from the parabolic reflector posses a planner wave front in the aperture plane, essential in producing the desired radiation pattern.

- Such a feed arrangement leads to a larger antenna noise temperature because the feed horn is pointed towards a relatively hot earth and therefore picks up a significant amount of noise. Additional between the feed and the low-noise amplifier (LNA); unless the LNA is mounted close to the waveguide used to connect the HPA (high power amplifier) to the antenna.

ii. Cassegrain and Gregorain systems

- The operating principle of a parabolic antenna is that a point source of radio waves at the focal point in front of a parabolic reflector will be reflected into a collimated plane wave beam along the axis of the reflector. Conversely, an incoming plane wave parallel to the axis will be focused to a point at the focal point.

- A typical parabolic antenna consists of a parabolic reflector with a small feed antenna at its focus, pointed back toward the reflector. The reflector is a metallic surface formed into a paraboloid of revolution and usually truncated in a circular rim that forms the diameter of the antenna.

- The reflector dish can be of sheet metal, metal screen, or wire grill construction, and it can be either circular or various other shapes to create different beam shapes. A mesh screen reflects radio waves as well as a solid metal surface as long as the holes are smaller than 1/10 of a wavelength, so screen reflectors are often used to reduce weight and wind loads on the dish.

- To achieve the maximum gain, it is necessary that the shape of the dish be accurate within a small fraction of a wavelength, to ensure the waves from different parts of the antenna arrive in phase. Large dishes often require a truss structure behind them to provide the required stiffness.

- The feed antenna at the reflector's focus is typically a low-gain type such as a half-wave dipole or more often a small horn antenna called a feed horn. In more complex designs, such as the Cassegrain and Gregorian, a secondary reflector is used to direct the energy into the parabolic reflector from a feed antenna located away from the primary focal point. The feed antenna is connected to the associated RF transmitting or receiving equipment by means of a coaxial cable transmission line or waveguide.

- *Feed pattern:* The radiation pattern of the feed antenna has a strong influence on the aperture efficiency, which determines the antenna gain (see next section). Radiation from the feed that falls outside the edge of the dish is called "spillover" and is wasted, reducing the gain and increasing the back lobes, possibly causing interference or (in receiving antennas) increasing susceptibility to ground noise. However, maximum gain is only achieved when the dish is uniformly illuminated with constant field strength to its edges. So the ideal radiation pattern of a feed antenna would be constant field strength throughout the solid angle of the dish, dropping abruptly to zero at the edges. However, practical feed antennas have radiation patterns that drop off gradually at the edges, so the feed antenna is a compromise between acceptably low spillover and adequate illumination.

iii. Asymmetric Configuration:

- The performance of an axi-symmetric configuration is affected by the blockage of the aperture by the feed and the sub-reflector assembly. The result is a reduction in the antenna efficiency and an increase in the side lobe levels. The asymmetric configuration can remove this limitation.

- This is achieved by offsetting the mounting arrangement of the feed so that it does not obstruct the main beam. As a result, the efficiency of the side lobe level performance is improved. The latter improvement is desirable because of the more stringent performance requirements. Hence the trend is to use this configuration where possible- especially for lower antenna sizes.

10.3.3) Antenna Mounts

- Description Today's requirements for secure, interoperable communications systems, as well as rapidly deployable networks for emergency response, are driving the need for inexpensive, simple, satellite earth station antennas, ranging in size from sub-meter to 5 meters in diameter. When selecting and siting antennas, systems engineers rarely consider the earth station antenna's vulnerability to damage or destruction by the forces of nature or man.

- Content when selecting the location for an earth station antenna of any size, the primary consideration is to ensure a clear view of the orbital arc, which allows the antenna to "see" the maximum number of satellites. Placing an antenna on a rooftop is often the optimal solution. However, extremely high winds can damage or destroy a parabolic dish antenna.

- Most properly installed earth station antennas are designed to survive winds of at least 60 or 70 miles per hour. When located in areas prone to hurricanes, tornadoes, or seasonal periods of high winds that can exceed these speeds, special considerations should be made in selecting the location of the antenna.

- Properly siting the antenna can increase the chances of surviving high wind conditions. Locating an antenna on the leeward side of buildings or hillsides, or using large roof structures, such as air conditioning units as windbreaks, while maintaining a clear view of the orbital arc, can make the difference between an antenna's survival or destruction in a storm. As every rooftop antenna installation is unique, it is important to work with the building owner or landlord in order to determine the optimum location.

- Many manufacturers make antennas and antenna mounts capable of surviving higher wind conditions, than standard units. High wind antennas are more robust, and reinforced mounts should be considered in areas having an elevated risk of wind damage to outdoor structures.

- In the case of non-penetrating roof mounted antennas being installed in high wind areas, the maximum amount of ballast recommended by the manufacturer must be installed, or even exceeded, in order to ensure that the antenna does not move from its moorings during high wind conditions. The ability of the roof to bear this additional load must be considered to avoid damaging the building upon which the antenna is mounted. In general, hard mounting an antenna to a building is preferred over the use of a non-penetrating roof mount.

- Having a replacement antenna available in the event of an emergency is a costly, yet highly effective means of mitigating the risk of prolonged outages in crucial networks. Installing a second, fully equipped and operational antenna on a nearby building provides full redundancy, and "pace diversity" for the system. While costly, this risk mitigating option may be appropriate in high priority, high value communications networks.

- **The most common antenna mounts used are:**
  - ***Azimuth elevation mount:*** An azimuth mount is a simple two-axis mount for supporting and rotating an instrument about two mutually perpendicular axes; one vertical and the other horizontal. Rotation about the vertical axis varies the azimuth (compass bearing) of the pointing direction of the instrument. Rotation about the horizontal axis varies the altitude (angle of elevation) of the pointing direction. These mounts are used, for example, with telescopes, cameras, radio antennas, heliostat mirrors, solar panels, and guns and similar weapons. Several names are given to this kind of mount, including altitude-azimuth, azimuth-elevation and various abbreviations thereof. A gun turret is essentially an alt-azimuth mount for a gun, with some armour to protect the weapon and its operators.

  - ***The X-Y Mount:*** This uses two orthogonal horizontal axes, one above the other, to permit dish rotation to any point by a combined rotation around each axis. Both designs suffer from the 'keyhole' problem. In the azimuth mount, the problem occurs at the zenith, that is, vertically above the azimuth axis. A satellite whose path passes close to the zenith will cause the antenna to slew rapidly as it tries to follow the spacecraft through a rapidly changing range of azimuth angles. Theoretically, the required slew rate approaches infinity as the satellite path approaches the zenith. With the advent of higher frequency operation, the antenna beam width is becoming narrower, and this leads to higher slew rate requirements if the signal is not to be lost near the zenith. For operation at 8 Ghz, the beam width is about 0.5 degrees. The solution has been to use large drive motors and bearings, with a consequent increase in the overall size and mass of the system, and a large intermittent load on the power supply.

    The X-Y mount has two 'keyholes' at opposite ends of a horizontal line through the pedestal, extended to the horizon. These horizon keyholes require a rapid exchange of angular position between the two axes, before the antenna can continue sweeping around or close to the horizon. Apart from the bearings and drives these mounts suffer the disadvantage of requiring large and costly pedestals. Along with heavy bearings and drives these are required to resist the substantial wind-age forces experienced on an inevitably exposed site. Again these forces are applied entirely through each bearing axis, which must be designed accordingly. Another problem with mounts of this type is that maintenance, for instance of a bearing, will often require the dish to be removed, which is an expensive undertaking.

10.4 EXERCISE

1. Explain the Azimuth Mount.
2. Discuss some general configurations of an Earth Station.
3. Write a note on Cassegrain and Gregorain systems.
4. Elaborate on X-Y mount antennas. How they are different from Azimuth mount antennas?
5. List and discuss the factors on which the general configurations of an earth station depend upon.

References

82. www.wikipedia.com
83. http://www.tech-faq.com/vsat.html
84. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
85. Rappaort, Wireless Communications Principals and Practices
86. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
87. P. Nicopolitidis ,Wireless Networks, John Wiley
88. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
89. http://www.britannica.com/EBchecked/topic/524891/satellite-communication

❖❖❖❖

11

EARTH STATIONS – PART II

Contents:

- Satellite television receivers

11.3 Exercise

11.1 GENERAL CONFIGURATION

11.1.1) Feed System

- The primary feed system used in existing earth stations performs a number of functions. Depending on the type of earth station, these functions may be:
  - To illuminate the main reflector.
  - To separate the transmit and receive bands to separate and combine polarizations in a dual polarized system.
  - To provide error signals for some types of satellite tracking system.

- A horn antenna is commonly used as the primary feed at microwave frequencies.A horn antenna consists of an open waveguide which is flared at the transmitting end so that the impedance of the free space matches the impedance of the waveguide.This ensures an efficient transfer of power.

- The figure below shows the block diagram of an orthogonal polarization feed assembly.A higher mode coupler (mode extractor) provides the error signal to the mono-pulse tracking system, if such a method is used.The orthogonal mode junction (OMJ) assembly is used to separate the dually polarized transmit and receive signal.
- The orthogonal mode transducer (OMT) separates the two linear orthogonally polarized signals into a composite linear orthogonally polarized signal on the transmit side.Because OMT operates on linearly polarized signals, polarizer's are used to convert a circular polarization to a linear.

- Polarizer's are therefore not required for linearly polarized system.Some earth stations have the capability to compensate polarization variations introduced by atmospheric effects by means of a feedback control system.The polarization properties of an antenna are mainly affected by the characteristics of the primary radiator and the polarizer.

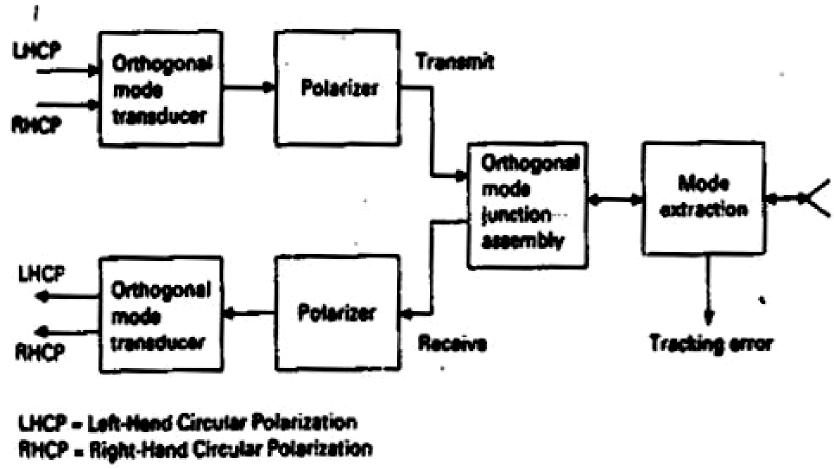


Figure 11.1: Feed System

11.1.2) Tracking System

- Tracking is essential when the satellite drift, as seen by an earth station antenna is a significant fraction of an earth station's antenna beam width.An earth station's tracking system is required to perform some of the functions such as
  - i) Satellite acquisition
  - ii) Automatic tracking
  - iii) Manual tracking
  - iv) Program tracking.

- The operation of the tracking system is explained by using its block diagram which show below.Communication satellites transmit a beacon which is used by earth stations for tracking.

- The received beacon signal is fed into the auto-track receiver where tracking corrections or, in some auto-track systems estimated positions of the satellite, are derived. In other auto-tack techniques the feed system provides the required components of error signal.
- The output of the auto-track receivers are processed and used to drive each axis of the antenna to the estimated satellite position.In manual mode, an operator sets the desired angles for each axis on a control console.

- This position is compared with the actual antenna position, obtained through shaft encoders, and the difference signal is used to drive the antenna.In the program track mode the desired antenna position is obtained form a computer.The difference in the actual and the desired antenna positions constitutes the error and is used to drive the antenna.

11.1.3) Recent Tracking Techniques

- There have been some interesting recent developments in auto-track techniques which can potentially provide high accuracies at a low cost. In one proposed technique the sequential lobing technique has been I implemented by using rapid electronic switching of a s single beam which effectively approximates simultaneous lobing.

- The high rate of switching is achieved by the use of an electronically controlled feed.This technique, sometimes referred to as electronic beam squinting, requires a simple single channel receiver and has been reported to achieve a tracking accuracy approaching that of the auto-track technique.

- Another approach, sometimes called as intelligent tracking, the satellite position is computed by optimal control techniques. The relatively complex computations are readily performed in an inexpensive microcomputer. The satellite position is obtained by optimally combining the antenna position estimate obtained from an accurate gradient tracking algorithm with predictions obtained from a simple, self learning satellite position model.

- There have also been some interest in employing phase array technique for satellite tracking specially in applications where the important design criteria are agility, low-profile and aesthetic. Here the antenna beam can be steered by exciting elements of an array antenna electronically. If a phase shift is introduced between successive elements of an array, the beam formed by the array is tilted in a direction determined by the sign of the phase-shift and the amount of tilt by its magnitude.

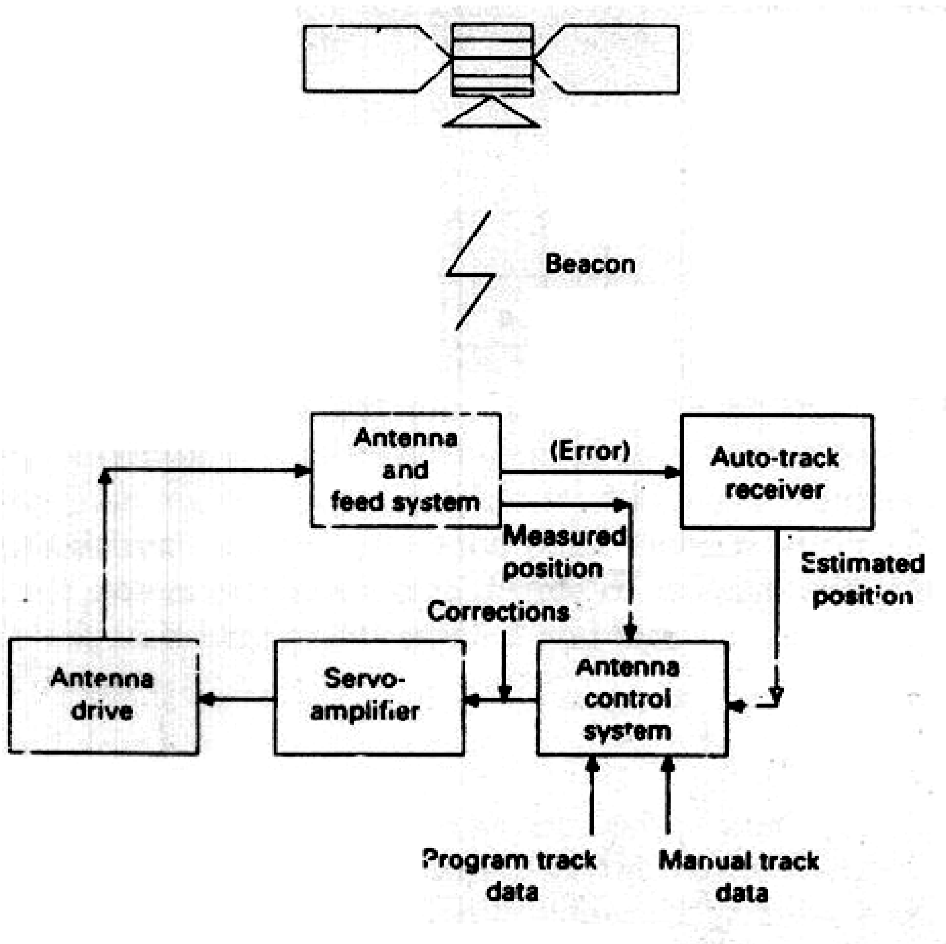


Figure 11.2: Tracking Device

11.1.4) Low noise amplifier

- In the earliest earth stations, MASERs were used as the front-end amplifier.These devices are relatively narrow band, require liquid helium temperatures and henceare expensive with difficult maintenance requirements.Thus, these were replaced by parametric amplifiers which could provide wide bandwidths, with the required low-noise temperatures at lower cost and complexity.

- Several improvements have been made to parametric amplifiers over the years.These have been made possible by the availability of improved devicesand the use of thermoelectric cooling. In recent years the advent of gallium arsenide field-effect transistors has greatly simplified the front-end amplifier design of earth stations.

- These devices provide similar orders of noise temperature and bandwidths as those of parametric amplifiers but at a lower cost.

11.1.5) High-power amplifier

- The high power amplifier (HPA) in an earth station provides the radio frequency (RF) carrier power to the input terminals of the antenna that, when it is combined with the antenna gain, it yields the equivalent isotropic radiated power (EIRP) required for the uplink to the satellite. The waveguide loss between the HPAis responsible for the calculation of the EIRP.

- The output power typically may be a few watts for a single data channel, around a hundred watts or less for a low capacity system or several kilowatts for high capacity traffic. The choice of amplifier is highly dependent on its application, the cost of installation and long term operation, and many other factors.

Types of amplifiers

- Generally, the earth station terminals use high power amplifiers designed primarily for operation in the Fixed Satellite Service (FSS) at C-band (6 GHz), military and scientific communications at X-band (8 GHz), fixed and mobile services at Ku-band (14 GHz), the Direct Broadcast Service (DBS) in the DBS portion of Ku-band (18 GHz), and military applications in the EHF/Q-band (45 GHz).

- Supplementary frequency bands include the ones allocated for the emerging broadband satellite services in Ka-band (30 GHz) and V-band (50 GHz). Mostly, the frequency used for the earth-to-space uplink is higher than the frequency for the space-to-earth downlink within a given band.

- An earth station HPA can be one of three types: a klystron power amplifier (KPA), a traveling wave tube amplifier (TWTA), or a solid state power amplifier (SSPA). The KPA and TWTA achieve amplification by modulating the flow of electrons through a vacuum tube.

- Solid state power amplifiers use gallium arsenide (GaAs) field effect transistors (FETs) that are configured using power combining techniques. The klystron is a narrowband, high power device, while TWTAs and SSPAs have wide bandwidths and operate over a range of low, medium, and high powers.

- The principal technical parameters characterizing an amplifier are its frequency, bandwidth, output power, gain, linearity, efficiency, and reliability. Size, weight, cabinet design, ease of maintenance, and safety are additional considerations. Cost factors include the cost of installation and the long term cost of ownership. KPAs are normally used for high power narrowband transmission to specific satellite transponders, typically for television program transmission and distribution.

- TWTAs and SSPAs are used for wideband applications or where frequency agility is required. Originally, TWTAs provided high power but with poor efficiency and reliability. Compared to a KPA, these disadvantages were regarded as necessary penalties for wide bandwidth. SSPAs first became available about 20 years ago. They were restricted to low power systems requiring only a few watts, such as small earth stations transmitting a few telephone channels.

11.2 CHARACTERISTICS

11.2.1) Fixed satellite service earth stations

- FSS earth stations use the C band, and the lower portions of the K_u bands. They are normally used for broadcast feeds to and from television networks and local affiliate stations as well as being used for distance learning by schools and universities, business television (BTV), Videoconferencing, and general commercial telecommunications. FSS satellites are also used to distribute national cable channels to cable television headends.Free-to-air satellite TV channels are also usually distributed on FSS satellites in the K_u band.

- In connection with the terrestrial networks, the traffic originating from these networks have to be reformatted according to the satellite network. This function is performed by an interface at the earth station, the configuration of which depends upon the type of traffic. Traffic signals may be available as frequency division multiplexed (FDM) analog telephony channels or time division multiplexed (TDM) streams with data or digitized telephony channels. Then traffic signals are demultiplexed at the earth station and rearranged on the basis of destination. A similar procedure is performed at the receiver's end.

- Fixed Satellite Service Earth Station uses ***VSAT*** is an abbreviation for a Very Small Aperture Terminal. It is basically a two-way satellite ground station with a less than 3 meters tall (most of them are about 0.75 m to 1.2 m tall) dish antenna stationed. The transmission rates of VSATs are usually from very low and up to 4 Mbit/s. These VSATs' primary job is accessing the satellites in the geosynchronous orbit and relaying data from terminals in earth to other terminals and hubs.

- They will often transmit narrowband data, such as the transactions of credit cards, polling, RFID (radio frequency identification) data, and SCADA (Supervisory Control and Data Acquisition), or broadband data, such as satellite Internet, VoIP, and videos. However, the VSAT technology is also used for various types of communications.

- Equatorial Communications first used the spread spectrum technology to commercialize the VSATs, which were at the time C band (6 GHz) receive only systems. This commercialization led to over 30,000 sales of the 60 cm antenna systems in the early 1980s. Equatorial Communications sold about 10,000 more units from 1984 to 1985 by developing a C band (4 and 6 GHz) two way system with 1 m x 0.5 m dimensions.

Implementations of VSAT

- Currently, the largest VSAT network consists of over 12,000 sites and is administered by Spacenet and MCI for the US Postal Service. Many huge car corporations such as Ford and General Motors also utilizes the VSAT technology, such as transmitting and receiving sales figures and orders, along with announcing international communications, service bulletins, and for distance learning courses.

- Two way satellite Internet providers also use the VSAT technology. Many broadband services around the world in rural areas where high speed Internet connections cannot be provided use VSAT.

VSAT Configurations

Most of the current VSAT networks use a topology:

- Star topology: This topology uses a central uplink site which transports the data to and from each of the VSAT terminals using satellites

- Mesh topology: In this configuration, each VSAT terminal will relay data over to another terminal through the satellite, acting as a hub, which also minimizes the need for an uplink site

- Star & Mesh topology: This combination can be achieved by having multiple centralized uplink sites connected together in a multi-star topology which is in a bigger mesh topology. This topology does not cost so much in maintaining the network while also lessening the amount of data that needs to be relayed through one or more central uplink sites in the network.

VSAT's Strengths

- VSAT technology has many advantages, which is the reason why it is used so widely today. One is availability. The service can basically be deployed anywhere around the world. Also, the VSAT is diverse in that it offers a completely independent wireless link from the local infrastructure, which is a good backup for potential disasters. Its deployability is also quite amazing as the VSAT services can be setup in a matter of minutes.

- The strength and the speed of the VSAT connection being homogenous anywhere within the boundaries is also a big plus. Not to forget, the connection is quite secure as they ar private layer-2 networks over the air. The pricing is also affordable, as the networks they do not have to pay a lot, as the broadcast download scheme allows them to serve the same content to thousands of locations at once without any additional costs. Last but not least, most of the VSAT systems today use onboard acceleration of protocols (eg. TCP, HTTP), which allows them to delivery high quality connections regardless of the latency.

VSAT's Drawbacks

- As the VSAT technology utilizes the satellites in geosynchronous orbit, it takes a minimum latency of about 500 milliseconds every trip around. Therefore, it is not the ideal technology to use with protocols that require a constant back and forth transmission, such as online games.

- Although VSAT is not as bad as one way TV systems like DirecTV and DISH Network, the VSAT still can have a dim signal, as it still relies on the antenna size, the transmitter's power, and the frequency band. Last but not least, although not that big of a concern, installation can be a problem as VSAT services require an outdoor antenna that has a clear view of the sky. An awkward roof, such as with skyscraper designs, can become problematic.

11.2.2) Mobile satellite service earth stations

- Satellites are well suited for a large area mobile communication. However the practical constraints imposed on the design of mobile earth stations meant that the constraints are imposed on the design of mobile earth stations means that the introduction of this service had to wait until the technology matured to a stage where these constraints could overcome in a cost efficient manner.

- The main features in design optimization of an SS earth station are:
  1) Limited mounting space implies that the antenna size on mobile is severely restricted.
  2) Minimization of the earth station cost is important for service uptake specially for personal communications.
  3) Traffic flow through the earth station is low.
  4) For personal communication the size, cost and power consumption approaches those of the cellular telephone, with a capability to operate with the part new terrestrial mobile system, with a capability to operate with a partner terrestrial mobile system. Transmitted power should conform to radiation safety standards.

11.2.3) Satellite Television Receivers

- It is atelevision delivered by the means of communications satellite and received by an outdoor antenna, usually a parabolic mirror generally referred to as a satellite dish, and as far as household usage is concerned, a satellite receiver either in the form of an external set-top box or a satellite tuner module built into a TV set.

- Satellite TV tuners are also available as a card or a USB stick to be attached to a personal computer. In many areas of the world satellite television provides a wide range of channels and services, often to areas that are not serviced by terrestrial or cable providers.

- Direct broadcast satellite television comes to the general public in two distinct flavors - analog and digital. This necessitates either having an analog satellite receiver or a digital satellite receiver.

- Analog satellite television is being replaced by digital satellite television and the latter is becoming available in a better quality known as high-definition television.

11.3 EXERCISE

1. Discuss the functions and features of VSAT.
2. Briefly discuss the general configurations of Earth Stations.
3. What are high power amplifiers?
4. Write a note on Tracking System. List its characteristics.
5. What are low noise amplifiers?
6. Discuss Fixed Satellite Service earth station.

References

90. www.wikipedia.com
91. http://www.tech-faq.com/vsat.html
92. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
93. Rappaort, Wireless Communications Principals and Practices
94. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
95. P. Nicopolitidis ,Wireless Networks, John Wiley
96. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
97. http://www.britannica.com/EBchecked/topic/524891/satellite-communication

❖❖❖❖

12

NON-GEOSTATIONARY ORBIT SATELLITE SYSTEMS

Contents:

12.1 INTRODUCTION

Non-geostationary satellites are mostly used for communication applications as the distance between them and earth is comparatively less and hence the delay seen is less. These satellites are also known as Low Earth Orbiting satellites (LEO) and Medium Earth Orbiting satellites (MEO).

12.2 REASONS

Although all types of orbits were considered for satellite communication during the initial years of satellite communication, geostationary orbits became very popular as they had several advantages, especially those pertaining to the antennas used for communicating with them.

The advantages and disadvantages of Non-geostationary satellites are given below:

12.2.1) Advantages

- LEO and MEO systems can provide true global coverage.
- Lower path loss makes it possible to use hand-held terminals.
- These satellites undergo low propagation delay.

12.2.2) Disadvantages

- A large number of satellites are required to cover the radius of the earth.
- The satellite visibility is 10-180 minutes which necessitates satellite-satellite handover leading to more complex network architecture.
- Doppler Effect is high.
- Signal strength at the receiver varies due to the continuous varying range and elevation angle.
- Maintenance of the network and the orbit in the longer term are challenging.
- These satellites also undergo a good number of eclipses; thus their batteries are expected to take charge of the satellite's functionality, hence their lifelong is also expected to be higher.
- Interferences of the satellite cannot be predicted as the distance is very high and the atmospheric effects are constantly changing.

12.3 DESIGN CONSDERATIONS

12.3.1) Traffic distribution and coverage

- Anorbit design is governed by the service area and geographical distribution of traffic within the area. In developing the coverage, distinction must be made between geometric visibility and RF visibility.

- A good RF visibility ensures that adequate signal strength is received before a connection is established. Further to increase the spectrum reuse, the coverage area is divided into smaller sets which are each covered by a spot beam.

- One of the unique traffic features of low earth orbit (LEO) satellite networks is time-variant and non-uniform load distribution. This feature results in a locally biased congestion problem for the LEO satellite systems. To solve the congestion problem, LEOs use of near-neighbor residual bandwidth information to allocate excess bandwidth from congested satellites to their under loaded neighbors in the network. Each traffic load balancing process is performed on the domain basis.

12.3.2) Satellite capacity

- LEO systems are designed to have more than one satellite in view from any spot on Earth at any given time, thus minimizing the possibility that the network will loose out on the transmission. LEO systems have to incorporate complicated tracking equipments to maintain consistent service coverage. The need for complex tracking schemes is minimized, but not obviated, in LEO systems designed to handle only short-burst transmissions.

- In addition, because the signals to and from the satellites need to travel a relatively short distance, LEOs can function with much smaller user equipment can systems using a higher orbit. In addition, a system of LEO satellites is designed to maximize the ability of ground equipment to "see" a satellite at any time, which can overcome the difficulties caused by obstructions such as trees and buildings.
  There are two types of LEO systems, Big LEOs and Little LEOs,

- **Little LEO** satellites are very small, often weighing no more than a human being, and use very little bandwidth for communications. Their size and bandwidth usage limits the amount of traffic the system can carry at any given time. However, such systems often employ mechanisms to maximize capacity, such as frequency reuse schemes and load delay tactics. Little LEO systems support services that require short messaging and occasional low-bandwidth data transport, such as paging, fleet tracking and remote monitoring of stationary monitors for everything from tracking geoplatonic movements to checking on vending machine status. The low bandwidth usage may allow a LEO system to provide more cost effective service for occasional-use applications than systems that maximize their value based on bulk usage.

- **Big LEO** systems are designed to carry voice traffic as well as data. They are the technology behind "satellite phones" or "global mobile personal communications system".Most Big LEO systems offer mobile data services and some system operators intend to offer semi-fixed voice and data services to areas that have little or no terrestrial telephony infrastructure. Smaller Big

LEO orbits also are planned to serve limited regions of the globe. MEO systems' larger capacity relative to LEOs may enable them to be more flexible in meeting shifting market demand for either voice or data services. MEO systems, as well as some Big LEOs, targeted at the voice communications market may have a disadvantage when compared with cellular and other terrestrial wireless networks. A satellite signal is inherently weaker and is more subject to interference than those of terrestrial systems, thus requiring a larger antenna than a traditional mobile phone. By contrast, the trend in the mobile phone market is toward smaller and smaller phones.

12.3.3) State of spacecraft technology

Orbit design is influenced by the chosen spacecraft technology. The following parameters are of significance.

- Antenna size and complexity: increasing the altitude trends to reduce the number of satellites in a orbit, but it also required the use of larger antennas to meet the link quality objectives and maintain the same frequency reusability.

- Spacecraft DC power: the DC power of a satellite determines the capacity of a satellite.

- Inter-satellite link: satellites with inter-satellite links influence the network routing scheme.

12.3.4) Terminal characteristics and communication requirement

- The size of the terminal and their communications capabilities influence a satellite's power and its sensitivity requirements. RF power of a handset is limited b safety considerations, battery size/capacity and the target terminal cost. If satellites are brought closer to a handset, power requirements reduce but the number of satellite in the orbit increases. Similarly, satellites require lower transmit power if the orbital altitude decreases.

12.3.5) Quality of service

- Quality of service (QoS) depends upon the reliability of RF link, propagation delays and signal quality measure as bit rate errors.

- For a given satellite EIRP; higher link reliability requires operation at higher elevation angle and path diversity ie more than one satellite must be visible from a terminal at a given time. Propagation conditions improve as the elevation angle increases because the number of obstructions reduces as the elevation angle is made to improve.

- Owing to the movement of satellite, propagation conditions fr LEOs and MEOs orbits are quite severe. The user could receive a high signal from a specific direction but not from others.
- Selection of the best orbit could be done from propagation and availability aspect can be done on the basis of the average, maximum and minimum elevation angles over the area of interest or using other criteria such as probability of a successful completion of voice calls.

12.3.6) Spectrum availability

- Each operation has access to a limited amount of RF spectrum and therefore whishes to maximize the use of this resource. Frequency reusability can be incremented by spatial and polarized diversity.

- Spatial reuse is achieved by spot and shaped beams. Within a specified area; and for a given spot beam size, a lower-altitude constellation tends to increase frequency reusability. Additional considerations in maximizing the radio recourse are modulation, coding and multiple access schemes.

12.3.7) Launch Considerations

- Important practical consideration relates to the launch cost, as well as the feasibility of launching the satellites in the constellation within an acceptable time frame. A number of launchers may be used to expedite launches as well as to spread the risk. The probability of launch failure and in-orbit satellite failure is likely to increase as the number of satellites in a constellation increases.

12.3.8) Orbital Debris

(As discussed in Unit 7)

- Besides planets, natural and artificial satellites, many other particles like cosmic rays, protons, electrons, meteoroids and manmade space debris exists in space.These particles collide with the satellites causing permanent damage to it and sometimes degrading the solar cells.

- Space debris, also known as orbital debris, space junk and space waste, is the collection of objects in orbit around Earth that were created by humans but no longer serve any useful purpose. These objects consist of everything from spent rocket stages and defunct satellites to explosion and collision fragments. The debris can include slag and dust from solid rocket motors, surface degradation products such as paint flakes, clusters of small needles, and objects released due to the impact of micrometeoroids or fairly small debris onto spacecraft. As the orbits of these objects often overlap the trajectories of spacecraft, debris is a potential collision risk.

- The vast majority of the estimated tens of millions of pieces of space debris are small particles, like paint flakes and solid rocket fuel slag. Impacts of these particles cause erosive damage, similar to sandblasting. The majority of this damage can be mitigated through the use of a technique originally developed to protect spacecraft from micrometeorites, by adding a thin layer of metal foil outside of the main spacecraft body.

- Impacts take place at such high velocities that the debris is vaporized when it collides with the foil, and the resulting plasma spreads out quickly enough that it does not cause serious damage to the inner wall. However, not all parts of a spacecraft may be protected in this manner, i.e. solar panels and optical devices (such as telescopes, or star trackers), and these components are subject to constant wear by debris and micrometeorites.

- The present means for spacecraft shielding, such as those used for the manned modules of the International Space Station are only capable of protecting against debris with diameters below about 1 centimeter. The only remaining means of protection would be to maneuver the spacecraft in order to avoid a collision. This, however, requires that the orbit of the respective object be precisely known.

- If a collision with larger debris does occur, many of the resulting fragments from the damaged spacecraft will also be in the 1 kilogram mass range, and these objects become an additional collision risk.

- As the chance of collision is a function of the number of objects in space, there is a critical density where the creation of new debris occurs faster than the various natural forces that remove these objects from orbit. Beyond this point a runaway chain reaction can occur that quickly reduces all objects in orbit to debris in a period of years or months.

12.3.9) Operational Considerations

- The monitoring and maintaining of a large orbit is a complex and expensive task. Monitoring has to be done in order to ensure that satellite transmissions meet the required specification in terms of frequency and power, traffic flow is normal, all satellites and links are operating satisfactorily, all the gateways are functioning correctly, reception is interference free and the user is satisfied with the signal quality.

- Maintenance of an orbit includes maintaining correct altitude and phase o satellites, which would typically require station-keeping every few weeks, replaced of failed spacecraft and so on. Again, the complexity tends to increase as the size of the orbit increases.

12.3.10) Network Issues

- Network issues comprise of network architecture including call connection strategy, intra and inter system routing considerations, mobility management and space segment resource management.

12.3.11) Network Architecture

- This depends up to the nature of services offered by the satellite, which is categorized as follows:
  - Non real time services such as messaging
  - Real time services such as voice or video conferencing
  - Combination of real time and non real time services.

- Non real time services include satellite based store and forward services; earth station based store and forward.

- Real time services include connectivity via inter-satellite links, distributed routing schemes, centralized routing schemes, flooding schemes, connection via ground relays, and connectivity via geostationary satellites.

12.3.12) Mobility Management

- In network, where user terminals are not fixed to a location, it becomes necessary to manage mobility of terminals so that the calls are successfully established. To minimize call set up time and network signaling requirements during the setup, and to improve the call set up success rate, it is necessary for the switching center t have knowledge of the location of each mobile in the network. The process of the terminal registering its location is called as location registration. The terminal or the network must have a provision for estimating the user's position.

- Handover increases the complexity of the network as the need for real-time signaling between various network entities. In general the number of handovers reduces with the increase in the orbital altitude because of a corresponding increase in the orbital period.

12.4 EXERCISE

1. Discuss the design considerations of Non-Geostationary orbit satellites.
2. Write a note on mobility management of Non-Geostationary orbit satellites.
3. How Non-Geostationary Orbit Satellites are different from Geostationary Orbit Satellites? Discuss their advantages and disadvantages.
4. Write a short note on Orbital Debris? How do they harm the Satellites?
5. Discuss the Satellite Capacity of Non-Geostationary orbit satellites.

References

98. www.wikipedia.com
99. http://www.tech-faq.com/vsat.html
100. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
101. Rappaort, Wireless Communications Principals and Practices
102. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
103. P. Nicopolitidis ,Wireless Networks, John Wiley
104. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
105. http://www.britannica.com/EBchecked/topic/524891/satellite-communication

❖❖❖❖

13

THE SPACE LINK – PART I

Contents:

13.1 INTRODUCTION

This unit describes how link-power budget calculations are done. These calculations generally relate two quantities, the transmission power and the receiver power. This unit also discusses how the difference between these two powers is accounted for.

Link-power budget calculations also need the additional losses and noise factor which is incorporated with the transmitted and the received signals. Losses can be of various types, the major ones considered for satellite communication are discussed here.

Along with losses, this unit also discusses the system noise parameters. Various components of the system add to the noise in the signal that has to be transmitted. Most of the calculations discussed in this unit are in decibel quantities.

13.2 EQUIVALENT ISOTROPIC RADIATED POWER

- The key parameter in link-power budget calculations is the equivalent isotropic radiated power factor, commonly denoted as EIRP. Is the amount of power that a theoretical isotropic antenna (which evenly distributes power in all directions) would emit to produce the peak power density observed in the direction of maximum antenna gain. EIRP can take into account the losses intransmission line and connectors and includes the gain of the antenna.

- The EIRP is often calculated in terms of decibels over a reference power emitted by an isotropic radiator with equivalent signal strength. The EIRP allows comparisons between different antennas in satellite communication regardless of type, size or form.EIRP can be defined as the power input to one end of the transmission link and the problem to find the power received at the other end.

EIRP = G Ps

Where,

G → Gain of the Transmitting antenna and G is in decibels.

Ps→ Power of the sender (transmitter) and is calculated in watts.

[EIRP] = [G] + [Ps] dBW

Exercise: A satellite downlink is at 12 GHz operate with a transmit power of 6 W and an antenna gain if 48.2 dB. Calculate the EIRP in dBW.

Solution:

[EIRP] = 10 log 6 + 48.2
= **56 dBW**

13.3 TRANSMISSION LOSSES

- As EIRP is thought of as power input of one end to the power received at the other, the problem here is to find the power which is received at the other end. Some losses that occur in the transmitting – receiving process are constant and their values can be pre – determined.

- Other losses can be estimated from statistical data and a few of them are dependent on the climatic conditions including rain and snow fall. To begin theses computations, generally the constant losses are determined considering a clear sky condition. Below listed are the losses which are generally taken as a constant value.

13.3.1) Free-Space Transmission Losses (FSL)

- This loss is due to the spreading of the signal in space. Going back to the power flux density equation (discussed in unit VI a):

$$\Psi_m = P_S / 4 \pi r^2$$

- The power that is delivered to a matched receiver is the power flux density. It is multiplied by the effective aperture of the receiving antenna. Hence, the received power is:

$$\begin{aligned} P_R &= \Psi_M A_{\text{eff}} \\ &= \frac{\text{EIRP}}{4\pi r^2} \; \frac{\lambda^2 G_R}{} \\ &= (\text{EIRP})\,(G_R)\left(\frac{\lambda}{4\pi r}\right)^2 \end{aligned}$$

Where
r→ distance between transmitter and receiver
G_R→ power gain at the receiver
In decibels, the above equation becomes:

$$[P_R] = [\text{EIRP}] + [G_R] - 10\log\left(\frac{4\pi r}{\lambda}\right)^2$$

$$[\text{FSL}] = 10\log\left(\frac{4\pi r}{\lambda}\right)^2$$

$$[P_R] = [\text{EIRP}] + [G_R] - [\text{FSL}]$$

13.3.2) Feeder Losses (RFL)

- This loss is due to the connection between the satellite receiver device and the receiver antenna is improper. Losses here occur is connecting wave guides, filers and couplers. The receiver feeder loss values are added to free space loss. Similar losses will occur in filters, couplers and waveguides that connect the transmission antenna to a high-power amplifier output.

13.3.3) Antenna Misalignment Losses (AML)

- To attain a good communication link, the earth station's antenna and the communicating satellite's antenna must face each other in such a way that the maximum gain is attained.

- Sometimes, misalignment (also called as off-axis loss) can occur in two ways:
  - The off-axis loss at satellite is taken into account by designing the link for operation on the actual satellite contour.
  - The off-axis loss at the earth station is referred to as antenna pointing loss. These losses are usually only a few tenths of a decibel.

- In addition to pointing losses, losses can occur due to the misalignment of the polarization direction. These losses are generally small and it will be assumed that the antenna misalignment loss includes pointing as well as polarization losses value.

- The value of this loss can be estimated using statistical data which are based on errors that are actually observed or a large number of earth stations.

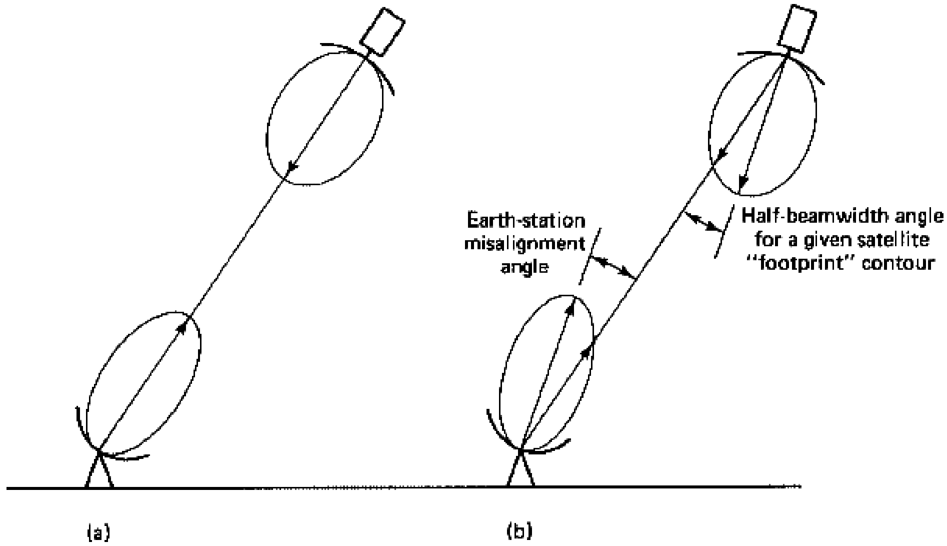


Figure 13.1: a) Satellite and Earth station's antennas aligned for maximum gain; b) Earth station is situated at the given satellite's footprint it's antenna is misaligned.

13.3.4) Fixed Atmospheric (AA) and Ionospheric losses (PL)
The gases present in the atmosphere absorb the signals. This kind of loss is usually of a fraction of decibel in quantity. Along with the absorption losses, the ionosphere introduces a good amount of depolarization of signal which results in loss of signal.

13.4) Link - Power Budget Equation

The EIRP can be considered as the input power to a transmission link. Due to the above discussed losses, the power at the receiver that is the output can be considered as a simple calculation of EIRP – losses.

$$\text{Losses} = [FSL] + [RFL] + [AML] + [AA] + [PL]$$

The received power that is P_R:$P_R = [EIRP] + [G_R] - [Losses]$
Where;
$[P_R]$→ received power in dB.
[EIRP] → equivalent isotropic radiated power in dBW.
$[G_R]$→ isotropic power gain at the receiver and its value is in dB.
[FSL]→free-space transmission loss in dB.
[RFL] → receiver feeder loss in dB.
[AA] →Atmospheric absorption loss in dB.
[AML] →Antenna misalignment loss in dB.
[PL] → depolarization loss in dB.

Example: a satellite link operating at 14 GHz has receiver feeder losses of 1.5 dB and a free-space loss of 207 dB. The atmosphericabsorption loss is 0.5 dB and the antenna pointing loss is 0.5 dB. Depolarization losses may be neglected. Calculate the total link loss for a clear – sky condition.

Solution: the total loss is the sum of all losses:

$$\text{Losses} = [FSL] + [RFL] + [AML] + [AA] + [PL]$$
$$= 207 + 1.5 + 0.5 + 0.5 + 0$$
$$\mathbf{= 209.5\ dB}$$

Where;
[FSL] → free-space transmission loss in dB.
[RFL] → receiver feeder loss in dB.
[AML] →Antenna misalignment loss in dB.
[AA] →Atmospheric absorption loss in dB.
[PL] → depolarization loss in dB.

13.5 SYSTEM NOISE

- Electrical noise is always present at the input and unless the signal is significantly larger than the noise, amplification will be of least help as it will amplify the signal as well as the noise to the same extent. There is a possibility, that after the amplification, the situation can get worst by the noise that will be added by the amplifier.

- The main source of noise in the satellite equipments is the noise arising from the random thermal motion of electrons in the various devices in the receiver. Thermal noise is also generated in the lossy components of the antenna and a thermal – like noise is picked – up by the antenna as radiation.
  Power from a thermal noise source is given by:
  $$P_N = k\, T_N\, B_N$$
  Where:
  T_N→ noise temperature
  B_N→ Noise Bandwidth
  k→Boltzman Constant having the value 1.38×10^{-23} J/k
  - The main characteristic of thermal noise is that it has a *flat frequency spectrum*; that is, noise power per unit bandwidth is a constant. The noise power per unit bandwidth is termed as *noise power spectral density*.
    Denoting this by NO
    $$N_O = P_N / B_N$$
    Thus, $\mathbf{N_O = k\, T_N}$Joules
  - Noise temperature is <u>directly proportional</u> to the physical temperature but not always equal.
  - Noise power per unit bandwidth is <u>always constant</u>.

Example: An antenna has noise temperature of 35 K and is matched into a receiver which has a noise temperature of 100 K calculate: a) noise power density and b) the noise power for a bandwidth of 36 MHz.
Solution:

a) $N_O = k\, T_N$
   $= 11.38 \times 10^{-23} \times (35 + 100)$
   $\mathbf{= 1.86 \times 10^{-21}\ J}$

b) $PN = N_O B_N$
$= 1.86 \times 10^{-21} \times 36 \times 10^{6}$
= 0.067 pW

13.5.1) Antenna Noise

- The received signal power is pointless unless compared with the power received from unwanted sources over the same bandwidth. Such noise sources consist of thermal radiation from the earth and sky, cosmic background radiation and random thermal processes in the receiving system. An additional noise due to non - stationary radio frequency interference from pagers, cellular phones, etc., often needs to be considered, but in this analysis we will concentrate on two classifications of the antennas noise: a) Sky noise, and, b) Noise originating from the antenna losses.

a) ***Sky Noise***: it is a term used to describe microwave radiation which is present throughout the universe and which appears to originate from matter in any form at finite temperature. Such radiation covers wider spectrum. Any absorptive loss mechanism generates thermal m\noise, there being direct connection between loss and the effective noise temperature. Rainfall introduces attenuation and thus it further degrades transmission in two ways:
1) It attenuates the signal; 2) it introduces noise.
The detrimental effects of rain are much worse at Ku-Band frequencies than at C-band (refer Unit I for Ku and C Band features and bandwidth), and the downlink rain fade margin also must allow the increased noise which is generated.

b) ***Antenna Losses:*** Satellite antennas are generally pointed towards the earth and therefore they receive the full thermal radiation from it. In this case the equivalent noise temperature of the antenna, excluding the antennas losses is approximately 290 K. Antenna losses add to the noise received as radiation and the total antenna noise temperature is the sum of equivalent noise temperatures of all these sources.

13.5.2) Antenna Noise Temperature

- Antenna noise temperature is the temperature of a theoretical resistor at the input of an ideal noise-free receiver that would generate the same output noise power per unit bandwidth as that at the antenna output at a specified frequency.Antenna noise temperature has contributions from several sources:
  - Vast radiation
  - Earth heating

- The sun
- Electrical devices
- The antenna itself

- The available power gain of the amplifier is denoted as G, and the noise power output as P_{no}. Considering noise power per unit bandwidth which is noise energy in joules is given by:

$$N_{o,ant} = k\, T_{ant}$$

- The output noise energy in $N_{o,out}$ will be $GN_{o,out}$ plus the contribution made by the amplifier. The summation of all the amplifier noise is referred to the input in terms of an equivalent input noise temperature for the amplifier Te. Thus output could be written as:

$$N_{o,out} = Gk\,(T_{ant} + T_e)$$

The total noise referred to the input is

$$N_{o,out}/\,G$$

OR

$$\mathbf{N_{o,in} = k\,(T_{ant} + T_e)}$$

13.5.3) Amplifier in Cascade

- A cascade amplifier is any amplifier constructed from a series of amplifiers, where each amplifier sends its output to the input of the next amplifier in a daisy chain.

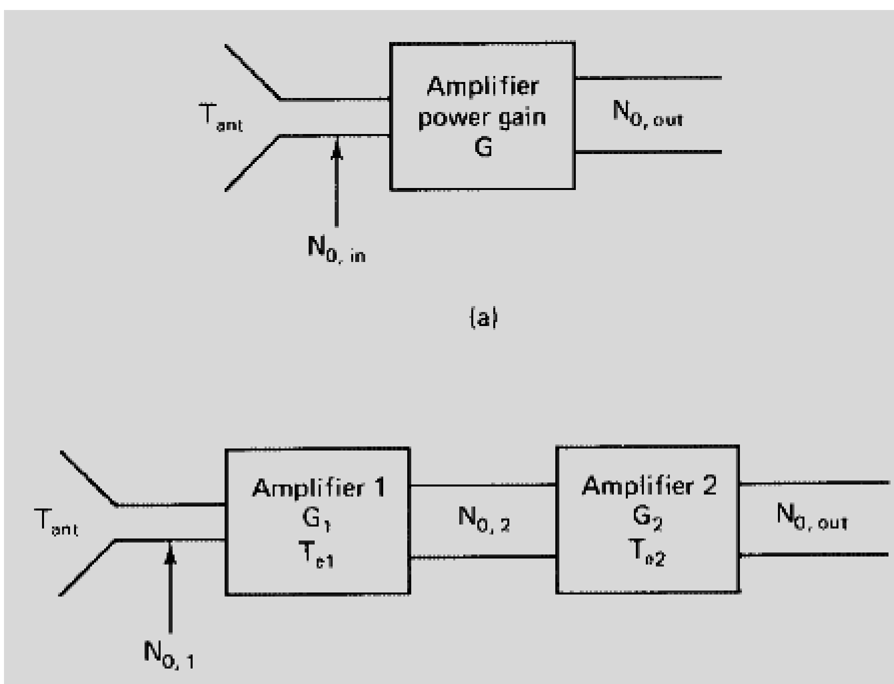


Figure 13.2: a) An amplifier; b) An amplifier in Cascade

For the arrangement of amplifiers shown in fig. 13.2 b; the overall gain can be considered as:

$$G = G_1\, G_2$$

- The noise energy of amplifier 2 referred to its own inputs is imply kT_{e2}. The noise input to amplifier 2 from the preceding stages is G_1 k (T_{ant} + T_{e1}), and thus the total noise energy referred to amplifier 2 input is:

$$N_{o,2}= G_1k\ (T_{ant} + T_{e1}) +kT_{e2}$$

- This noise energy may be referred to amplifier 1 input by dividing by the available over gain of amplifier 1:

$$N_{o,2}= N_{o,2} / G_1$$
$$= k\ (T_{ant} + T_{e1} + T_{e2} / G_1$$

A system noise temperature may now be defined as T_S by

$$N_{o,1}= k\ T_S$$

And hence it will be seen that $T_S$is given by

$$\mathbf{T_S = T_{ant} + T_{e1}+ T_{e2} / G_1}$$

13.5.4) Noise Factor

- Definition: An alternative way of representing amplifier noise is by the means of its noise factor F. for defining it, the source is taken at room temperature, denoted by T_0. The input noise from such a source is kT_0 and the output noise from the amplifier is:

$$N_{o,out}= FGkT_0$$

Where:

G is the available power gain of the amplifier
F is its noise factor

13.5.5) Noise Temperature of Absorptive Networks

- An absorptive network is one which contains resistive elements. These introduce losses by absorbing energy from the signal and converting it into heat. Resistive attenuators, transmission lines and wave guides are all examples of absorptive networks. Even natural phenomenon like rainfall, which absorbs energy from radio signals passing through it can be considered as a form of absorptive network. As these absorptive networks contain resistance, they generate thermal noise.

13.5.5) Overall System Noise Temperature

It's a summation of all the above discussed noise parameters. It is denoted as T_S. This parameter of system noise is considered for satellite communication computations.

13.6 SUMMARY

This unit discusses the equivalent isotropic radiated power factor which can be defined as the power input to one end of the transmission link and the problem to find the power received at the other end.

Further, this unit discusses various transmission losses face by a signal travelling between earth station and satellite. They include atmospheric losses, depolarization losses, misalignment of antenna loss, feeder loss and free space loss. Combining these losses, one can calculate the link-power budget value and estimate the loss of a particular signal by studying the statistics of a communication link.

Further, this unit discusses the system noise faced in a satellite. This includes atmospheric loss which is calculated with the antenna loss, sky noise, noise temperature of an antenna and an amplifier cascade. Along with these noise parameters, noise factor are also calculated which helps in calculating the overall system noise of a spacecraft.

13.7 EXERCISE

1. Write a note on Equivalent Isotropic Radiated Power (EIRP).
2. Discusses the transmission losses seen between a space craft and earth station.
3. What is antenna misalignment loss? Propose a solution to overcome these losses.
4. Derive the Link-Power Budget equation.
5. Write a note on system noise.

References

106. www.wikipedia.com
107. http://www.tech-faq.com/vsat.html
108. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
109. Rappaort, Wireless Communications Principals and Practices
110. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
111. P. Nicopolitidis ,Wireless Networks, John Wiley
112. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
113. http://www.britannica.com/EBchecked/topic/524891/satellite-communication

❖❖❖❖

14

THE SPACE LINK – PART II

Contents:

14.1 INTRODUCTION

This unit discusses the Noise factor faced by the signal while transmission. Considering the difficulties achieving a terrestrial link, it might be surprising that satellitelinks, covering much greater distance are possible at all. One of the most important factorsto explain this is the noise involved.When signals originate at the satellite; they are virtuallyfree of noise. The origin of noise and the meaning of the noise pressure and temperature inrelation to receivers will be explained. The effect of the Earth atmosphere on signal-to-noiseratio (SNR) will be illustrated with a real-life example.This noise is considered separately for uplink and downlink. Based on the transmission / receiving frequency band, a carrier to noise ratio for every travelling signal is calculated to determine the loss in each signal. Once the noise value is estimated, measures to overcome this are calculated.

14.2 NOISE

- Many types of noise are transmission related. Sometimes it's nothing more than a normal noise that sounds louder because of bad bases or because part of the transmission is touching the frame or underbody of the car. Then there are actual components like pumps, planets, final drives etc. that can cause good amount of disturbance in any travelling wave. The idea is to find a way to make the noise change, or stop, and then examine what this change did to affect the noise.

- There are several rules that will help isolate the component that is causing the problem. First of all, a component cannot generate a noise if it is not moving.Isolating moving components and calculating statistically the amount of noise they produce can help us estimate the signal loss.

- Next, if the noise is pressure related, it will change when the pressure changes.So again, estimation helps in determining the loss that can occur in a particular signal. There is variation in n noise in a particular link while the signal is moving upward and while is it moving downward. Presence and absence of atmospheric pressure, gravity and amount/ impact of sun's radiation also add to the noise factor of a signal.

14.3 CARRIER – TO – NOISE RATIO

A measure of a performance of a satellite link is considered as a ratio of carrier power to noise power at the receiver input along with the link budget calculations which are considered to estimate this ratio. This ratio is denotes as C/N and is calculated in decibels.

$$C/N = [P_R] – [P_N]$$

where:
C/N → carrier to noise ratio
P_R→ Receiver Power
P_N→Noise Power

Thus, the resultant C/N can be calculated with the following parameters (for parameters refer unit 13):

$$[C/N] = [EIRP] – [G_R] – [LOSSES] – [k] – [T_S] – [B_N]$$

To complete the calculations, we need to consider the gain is to temperature ratio as well. It is commonly denoted ad G/T. it is denoted as:

$$[G/T] = [G_R] – [T_S]$$

Thus, the C/N equation could be written as:

$$\mathbf{[C/N] = [EIRP] + [G/T] - [LOSSES] – [k] – [B_N]}$$

The ratio of carrier to noise power density P_R / $P_N$can be the quantity that is actually required. Since $P_N = k\ T_N B_N$ (proved in Unit 13), then:

$$[C/N] = [C / N_O\ B_N]$$
$$= [C / N_O] – [B_N]$$

And therefore,

$$[C/No] = [C / N_O] +[B_N]$$

[C/N] is true power ration in units of decibels, and $[B_N]$ is in decibels relative to one hertz or dBHz. Thus the units for [C/No] are dBHz. Applying this value to the above equation, we get:

$$[C/N_0] = [EIRP] + [G/T] - [LOSSES] - [k]$$

14.4 THE UPLINK

The uplink of a satellite circuit is where the earth station is transmitting the data to the space craft and the space craft is receiving it. The above discusses carrier – to – noise ratio equation can be determined for an uplink with an annotation of U. it is given by:

$$[C/N_0]_U = [EIRP]_U + [G/T]_U - [LOSSES]_U - [k]$$

14.5 THE DOWNLINK

The downlink of a satellite circuit is where the space craft is transmitting the data to the earth station and the earth station is receiving it. The above discusses carrier – to – noise ratio equation can be determined for andownlink with an annotation of D. it is given by:

$$[C/N_0]_D = [EIRP]_D + [G/T]_D - [LOSSES]_D - [k]$$

14.6 COMBINED UPLINK AND DOWNLINK C/N RATIO

- The complete satellite circuit consists of an uplink and a downlink. Noise will be introduced on the uplink at the satellite receiver input. Denoting the noise power per unit bandwidth by P_{NU} and the average carrier at the same point by P_{RU}, the carrier – to – noise ratio on the uplink is:

  $$(C/N_0) / U = (P_{RU} / P_{NU})$$

- It is important to note that power levels, and not decibels, are being used here. The carrier power at the end of the space link is shown as P_R, which of course is also the received carrier power for the downlink. This is equal to γ times the carrier power input to earth station input, as given in the below figure.

- It includes the satellite transponder and transmits antenna gains, the downlink losses, and the earth station receives antenna gain and feeder losses. The noise at the satellite input also appears at the earth station input multiplied by γ and in addition, the earth station introduces its own noise which is denoted as P_{ND}. Thus the end – of – link noise is $\gamma P_{NU} + P_{ND}$.

- The C/N_0ratio for the downlink alone, not counting the γP_{NU}contribution, is $P_R / P_{ND}$and the combined C/N_0ratio at the ground receiver is $P_R(\gamma P_{NU} + P_{ND})$. The power flow diagram is shown figure 14.1 b). The combined carrier – to – noise ratio can be determined in terms of the individual link values. To show this, it is more convenient to work with the noise – to – carrier ratios rather than the carrier – to – noise ratio, and again, these must be expressed as power ratios, not decibels. Denoting the combined carrier – to – noise values by N_0 / C, the uplink value by $(N_0 / C)_U$and the downlink value by $(N_0 / C)_D$ then,

$$N_0 / C = P_N / P_R$$
$$= (\gamma P_{NU} + P_{ND})/ P_R$$
$$= (\gamma P_{NU}/ P_R) + (P_{ND}/ P_R)$$
$$= (\gamma P_{NU}/\gamma P_R) + (P_{ND}/ P_R)$$
$$\mathbf{N_0 / C = (N_0 / C)_U + (N_0 / C)_D}$$

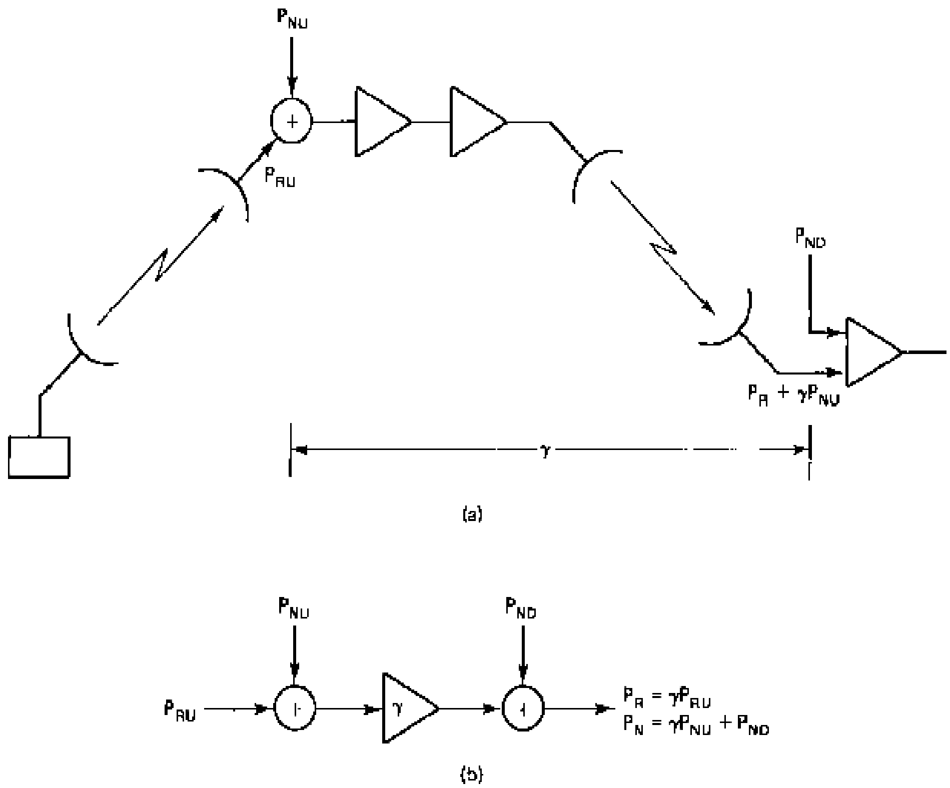


Figure 14.1: a) Combined uplink and downlink; b) power flow diagram for (a)

The above derived equation is the combine value of C/N_0 , the reciprocals of individual values must be added to obtain the N_0 / C ratio and then reciprocal of this taken to get C/N_0.

14.7) INTERMODULATION NOISE

- Intermodulation noise is the amplitude modulation of signals containing two or more different frequencies in a system with nonlinearities. The intermodulation between each frequency component will form additional signals at frequencies that are not just atharmonic frequencies (integer multiples) of either, but also at the sum and difference frequencies of the original frequencies and at multiples of those sum and difference frequencies.

- Intermodulation is caused by non-linear behavior of the signal processing being used. The theoretical outcome of these non-linearities can becalculated by generating a series of the characteristic, while the usual approximation of those non-linearities is obtained by generating aTaylor series.

- Intermodulation is rarely desirable in radio or audio processing, as it creates unwanted spurious emissions, often in the form of sidebands. For radio transmissions this increases the occupied bandwidth, leading to adjacent channel interference, which can reduce audio clarity or increase spectrum usage.

- In satellite communication systems, this most commonly occurs in the travelling – wave tube high – power amplifier abroad the satellite. Both amplitude and phase non – linearities give rise to this intermodulation noise.

- The carrier – to – intermodulation - noise ratio is usually found experimentally or in some cases it may be determined by computer methods. Ratio can be combined with the carrier –to – thermal noise ratio by the addition of the reciprocals. Denoting intermodulation noise term by $(C/N)_{IM}$ and bearing in mind that the reciprocals of C/N power ratios must be added. The ratio can be re-written as:

$$(N_0/C)= (N_0/C)_U+(N_0/C)_D+ (N_0/C)_{IM}$$

14.8 SUMMARY

This unit discusses the Noise in satellite commination link. This noise is calculated using two forms: Uplink and Downlink. The uplink of a satellite circuit is where the earth station is transmitting the data to the space craft and the space craft is receiving it. The downlink of a satellite circuit is where the space craft is transmitting the data to the earth station and the earth station is receiving it.

Further it derives an equation for combined uplink and downlink which is:

$N_0 / C = (N_0 / C)_U + (N_0 / C)_D$.

The last segment of the unit discusses Intermodulation noise. Intermodulation noise is the amplitude modulation of signals containing two or more different frequencies in a system with nonlinearities. This noise is commonly seen in satellite communication link. To conclude the unit, the total noise is computed combining the uplink, downlink and the intermodulation noise and is given by the equation: $(N_0/C) = (N_0/C)_U + (N_0/C)_D + (N_0/C)_{IM}$

14.9 EXERCISE

6. Write a note on system noise.
7. Diagrammatically explain the combined Uplink and Downlink carrier to noise ratio.
8. What is understood by intermodulation noise?
9. Derive the Equation for Combined Uplink and Downlink.
10. What is Uplink?
11. What is Downlink?

References

114. www.wikipedia.com
115. http://www.tech-faq.com/vsat.html
116. M. Richharia, Mobile Satellite Communication: Principles and Trends, Pearson Education
117. Rappaort, Wireless Communications Principals and Practices
118. YI Bing Lin, Wireless and Mobile Network Architectures, John Wiley
119. P. Nicopolitidis ,Wireless Networks, John Wiley
120. Satellite Communications Dennis Roddy 3rd edition, Mc-Graw Hill publication
121. http://www.britannica.com/EBchecked/topic/524891/satellite-communication

❖❖❖❖

CLASS: M. Sc (Computer Science)
Sub: Satellite Communications

Introduction:

General background, frequency allocations for satellite services, basic satellite system, system design considerations, applications.

2. Satellite Orbits:

Introduction, laws governing satellite motion, orbital parameters, orbital perturbations, Doppler effects, geostationary orbit, antenna look angles, antenna mount, limits of visibility, Earth eclipse of satellite, sun transit outage, inclined orbits, sun-synchronous orbit, launching of geostationary satellites.

3. Wave Propagation and Polarization:

Introduction, atmospheric losses, ionospheric effects, rain attenuation, other impairments, antenna polarization, polarization of satellite signals, cross polarization discrimination, ionospheric depolarization, rain depolarization, ice depolarization.

4. Satellite Antenna:

Antenna basics, aperture antennas, parabolic reflectors, offset feed, double reflector antennas, shaped reflector systems.

5. Link Design:

Introduction, transmission losses, link power budget equation, system noise, carrier to noise ratio for uplink and downlink, combined uplink and downlink carrier to noise ratio, inter modulation noise

6. Communication Satellites:

Introduction, design considerations, lifetime and reliability, spacecraft sub systems, spacecraft mass and power estimations, space segment cost estimates.

7. Earth Stations:

Introduction, design considerations, general configuration and characteristics.

8. Multiple Access Techniques:

Introduction, FDMA, TDMA, FDMA/TDMA, operation in a multiple beam environment, CDMA, multiple access examples

9. Non Geostationary Orbit Satellite Systems:

Introduction, reasons, design considerations, case study, example of systems.

❖❖❖❖