Historical Background :—

⑧

* The first communication from an artificial earth satellite took place in october 1957 when the Soviet satellite Sputnik I, transmitted telemetry information for 21 days.

* After this Explorer I, launched in January 1958 by the United states & transmitted telemetry for nearly five months.

* The first artificial satellite used for voice communication was Score, launched in Dec 1958 and used to broadcast President Eisenhower's christmas message of that year.

* In those early years, serious limitations were imposed on payload size by the capacity of launch vehicles and the reliability of space borne electronics. To solve some of those problems, an experimental passive repeater, Echo-I was placed in medium altitude orbit in 1960. Signals were reflected from the metallized surface of this satellite, which was simply a large ballon. The approach was

Simple and reliable, but huge transmitters were needed on the ground to transmit even very low rate data.

* During the same year (1960) Courier, a store and forward satellite that put messages on magnetic tape for retransmission later during the orbit.

* The first nongovernment ventures into space communications occured in July 1962, when the Bell System designed and built Telstar I, an active real time repeater. Telstar was placed in a medium altitude elliptical orbit by NASA.

* The government experiments continued, with NASA launching Relay I in Dec 1962. This satellite, built by RCA (Radio Corporation of America), was used for early experiments with the transmission of voice, video and data.

* Perhaps the most important questions considered in the early 1960s centered around the best orbit to use for a communication satellite. Medium altitude systems have the advantage of low launch costs, higher payloads, and relatively short radio frequency propagation times. Their

disadvantage is the need to track the satellite in orbit with tracking earth stations and to transfer operations from one satellite to another. no single satellite link is available at all times for all stations in the network. The use of geostationary orbit was first suggested by Arthur C. clarke in the mid 1940S. The advantage of this orbit is that nearly the whole earth can be covered with three satellites, each maintaining a stationary position and able to see one third of the earths surface. No hand over is needed

* The 1st attempt at a synchronous orbit was made by NASA, launching SYNCOM I in feb 1963. Although SYNCOM I was lost at the point of orbit injection, SYNCOM II and SYNCOM III, launched in July 1963 & Aug 1964 respectively were able to accomplish successful synchronous orbit placement.

* The communications satellite Act allowed for the formation of the communications satellite corporation (Comsat) and provided the environment to spawn one of the most successful multinational ventures ever undertaken, Intelsat.

* Intelsat was formed in July 1964 and made the courageous decision to "go synchronous" and launched Early Bird (Intelsat I) in April 1965 into that orbit. It was a milestone in the development of satellite communications for commercial use.

* The series from Intelsat I through Intelsat IVA were successively larger spin stabilized ~~or~~ spacecraft

* Russia has a series of geostationary satellites to provide fixed, mobile and broadcast services in that country and in addition, starting in 1965 has used Molniya satellites in a highly elliptical orbit to provide TV and voice distribution.

Basic concepts of satellite communications:-

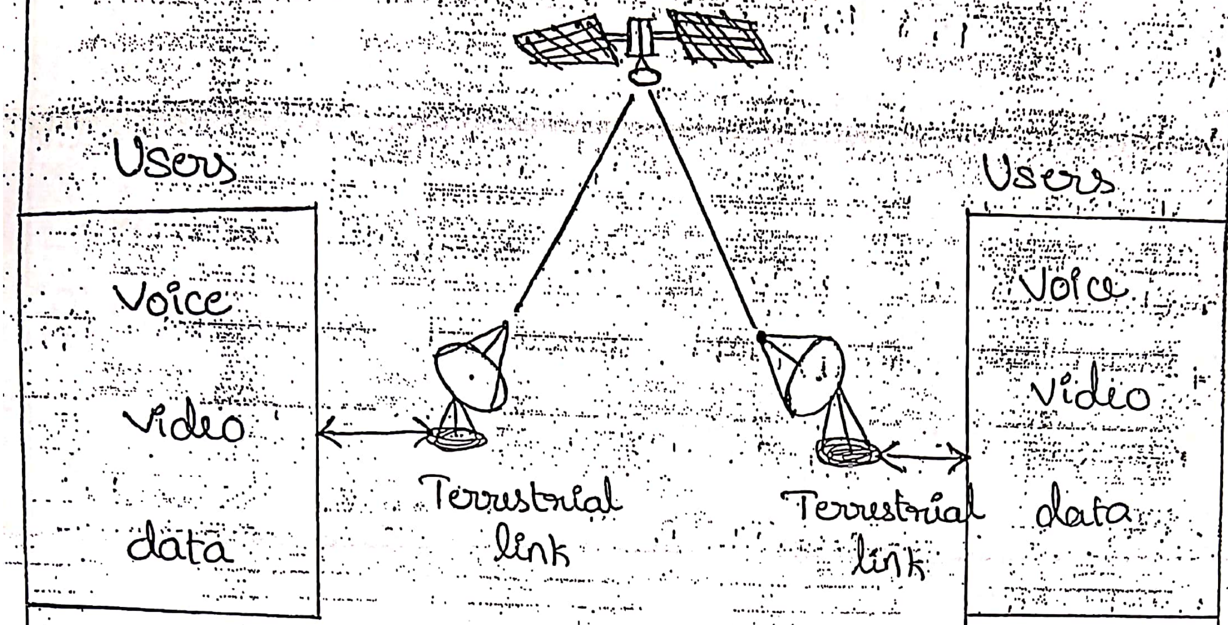


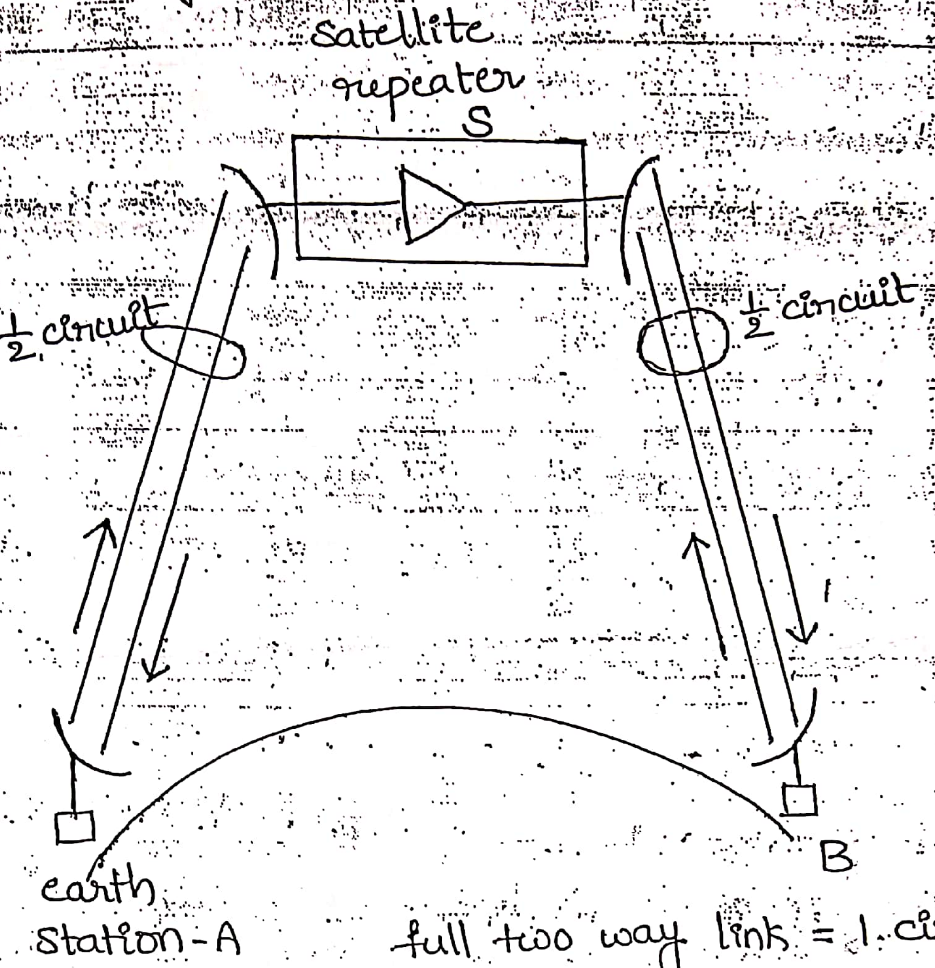
fig: General satellite link

Regardless of the orbits used and the communications services provided all the satellite links have some elements in common. The above figure illustrates the end to end communications required in establishing a satellite link. The link is shown in its most general form with transmit and receive facilities at both ends. Such facilities are characteristic of the fixed and mobile services but broadcast and data collection applications are transmitted only at one end and received only at the other end of the link. The overall problem can be conveniently divided into two parts.

The first deals with the satellite radio frequency (RF) link, which establishes communications between a transmitter and a receiver using the satellite as a repeater. In describing the satellite radio link, we quantify its capability in terms of the overall available carrier to noise ratio $(\frac{C}{N})$. The value of $(\frac{C}{N})$ depends on a no.of factors, which in turn depend on the available power and bandwidth.

The second part concentrates on the link between the earth terminal and the user environment. In the user environment, customers are typically concerned with establishing voice, data or video communications with either simplex or duplex connections. The quality of these baseband links is characterized by various figure of merit such as transmission rates, error rate, signal to noise ratio and other performance measures. (For example financial accounts link should have low error rate) The two parts of the problem can then be linked together when the available $\frac{C}{N}$ ratio of the satellite link is compared to the required $\frac{C}{N}$ ratio dictated by the user.

Radio frequency satellite link :-



full two way link = 1 circuit

one way link (A to B) = 1 channel

Two way link (A to S) = $\frac{1}{2}$ circuit.

fig : Basic concept of satellite communications

A communication satellite operates as a distant line of sight microwave repeater providing communications services among multiple earth stations in various geographic locations. The performance of a satellite link is typically specified in terms of its channel capacity.

* A channel is a one way link from a transmi earth station through the satellite to the receiving earth station

* A circuit is a full duplex link between two earth stations

* A half circuit is a two way link between an earth station and the satellite only.

The capacity of a link is specified by the types and no of channels and the performance requirements of each channel. In the case of international systems, a link from a transmitting station to the satellite may originate in one country and the link from the satellite to the receiving earth station may terminate in a second country. In this case, the concept of a half circuit is used for accounting purposes.

The channel carrying capacity of a satellite RF link is directly related to the overall available carrier to noise ratio. The carrier to noise ratio at the earth station, on which the performance depends, is the ratio of the carrier received from the satellite on the downlink to the total noise at the earth station from all sources. This noise comprises principally the thermal and receiver device noise at the earth station, to which must be added the satellite receiver noise retransmitted on the downlink and atmospheric and cosmic noise received at the antennas. The system designer must compromise among these three considerations: power, bandwidth, interference. The technical trade offs are often difficult and complicated by economic and regulatory factors.

The second component in the RF link is the downlink carrier to noise ratio $\left(\frac{C}{N}\right)_D$. As with the uplink, $\left(\frac{C}{N}\right)_D$ depends on the power of the transmitter, the transmitting and receiving antenna gains and the receiving system noise temperature. The third component to be considered in the RF link design is the satellite electronics system itself, which produces undesirable noise like signals that are normally expressed in a $\frac{C}{N}$ ratio that we shall call $\left(\frac{C}{N}\right)_I$

Satellite transponders

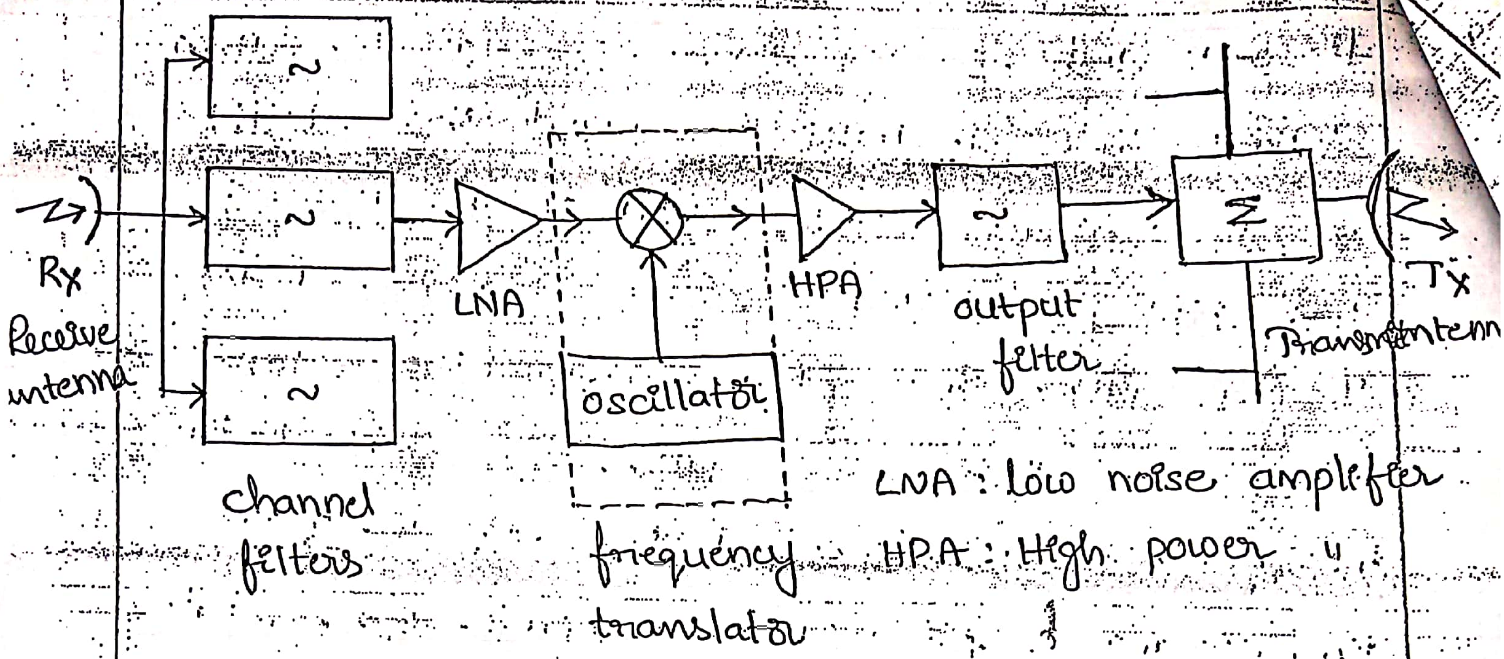


fig : Basic satellite repeater

From a communications standpoint, a satellite may be considered as a distant microwave repeater that receives uplink transmissions and provides filtering ; amplification, processing and frequency translation to the downlink band for retransmission. The kind of transponder is a quasi linear repeater amplifier, a block diagram of which is shown in figure above. The uplink and downlink bands are separated in frequency to prevent oscillation within the satellite amplifier, while permitting simultaneous transmission and reception at different frequencies through a device called the multiplexer

Satellite transponder amplifiers must provide large gains, while maintaining low noise operation. The high gain requirements typically require multiple stage low noise amplification. The first stages in modern transponder amplifier chains are provided by solid state FET amplifiers. These devices require careful design to minimize noise and intermodulation effects. Channelizing filters must also employ careful design to minimize interference from adjacent channels, as well as intersymbol interference and group delay distortion. Final stages of amplification in the transponder are typically provided by travelling wave tube amplifiers (TWTAs), which operate well for constant envelope signals. In the high power output amplifier stage that most of the impairements that affect $\left(\frac{C}{N}\right)$ are generated. These impairements are related to both the design of the satellite components and the operating points in the RF link.

Earth stations :

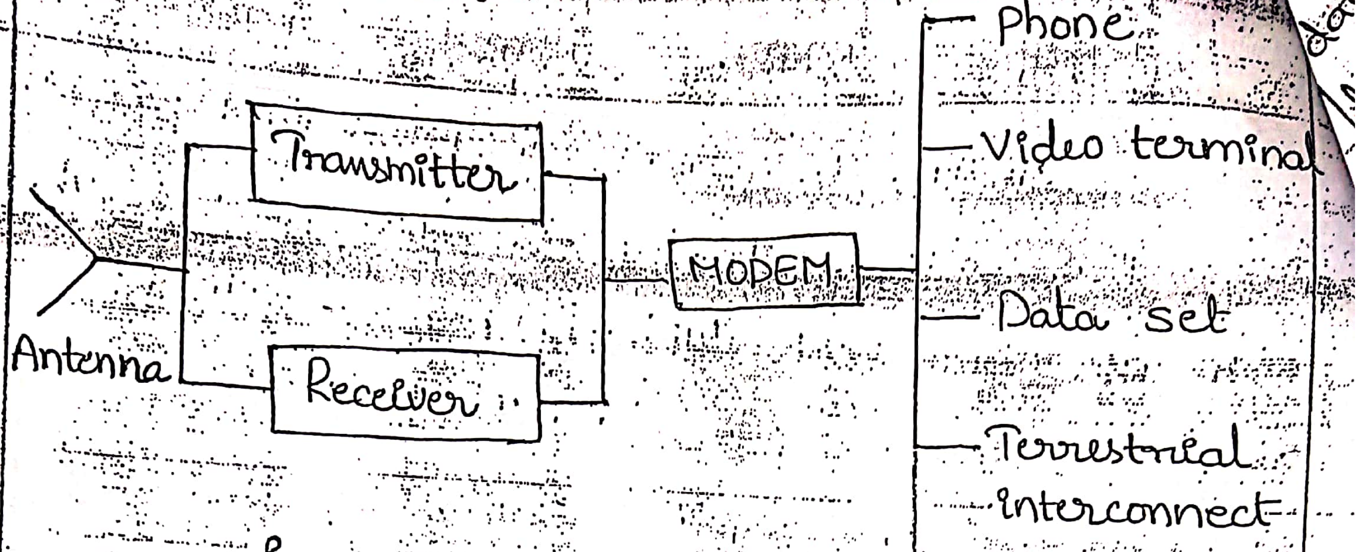


fig: Generic earth terminal

In the early days of satellite communication, most earth stations were both transmit and receive with large antennas, frequently in excess of 30m, transmit powers in excess of 5KW and cryogenic receivers with noise temperatures around 20K.

When we leave the domain of transmit and receive stations and consider receive only stations, such as used in cable installations to receive TV for redistribution by cable, TVRO stations for the direct reception of video from satellites and direct reception of audio, navigational, and still other kinds of electronic information, the variety of earth terminals becomes substantial. Finally we have the developing category of transmit only stations for the satellite reception and retransmission of data and messages, either in realtime or in store and forward mode.

A block diagram of a typical earth station is shown in figure. Earth stations are available in a wide variety of size, function, sophistication, and cost. They are categorized by function, by the size of the antenna, and by the level of the radiated power.

An earth station consists of an antenna subsystem, a power amplifier subsystem, a low noise receiver subsystem and a ground communications equipment (GCE) subsystem. Most stations are equipped with separate power supply systems, plus control, test, and monitoring facilities, sometimes called telemetry, tracking and command systems (TT&C).

The performance of an earth station is specified by its equivalent isotropic radiated power (EIRP) and its gain to system noise temperature ratio $\left(\frac{G}{T}\right)$. EIRP is the product of the power output of the high power amplifier at the antenna and the gain of the transmitting antenna. The receiving system sensitivity is specified by $\frac{G}{T}$, the ratio of the receive gain of the antenna to the system noise temperature.

The antenna gain is proportional to the square of the diameter and is dependent on the efficiency of the feed/reflector system. The system noise temperature is composed of three components : The noise of the receiver, the noise due to losses between the antenna feed system and the receiver, and the antenna noise. Although the performance of an earth station is typically limited by thermal noise, it can also be plagued with some of the same difficulties caused by nonlinear impairements in a satellite transponder.

The terrestrial link :

The 2nd part of the end to end satellite communications problem is embedded in the link between the satellite earth station and the user environment. This part deals more specifically with the baseband signal. To provide adequate satellite service to a user, the service requirements must be well defined in terms of quality. Quality of service specified in terms of parameters such as link availability (grade of service), bit error rate, and signal to noise ratio (carrier to noise ratio). The required $\frac{C}{N}$ ratio is then compared with the available $\frac{C}{N}$ ratio to determine the overall capacity of the link.

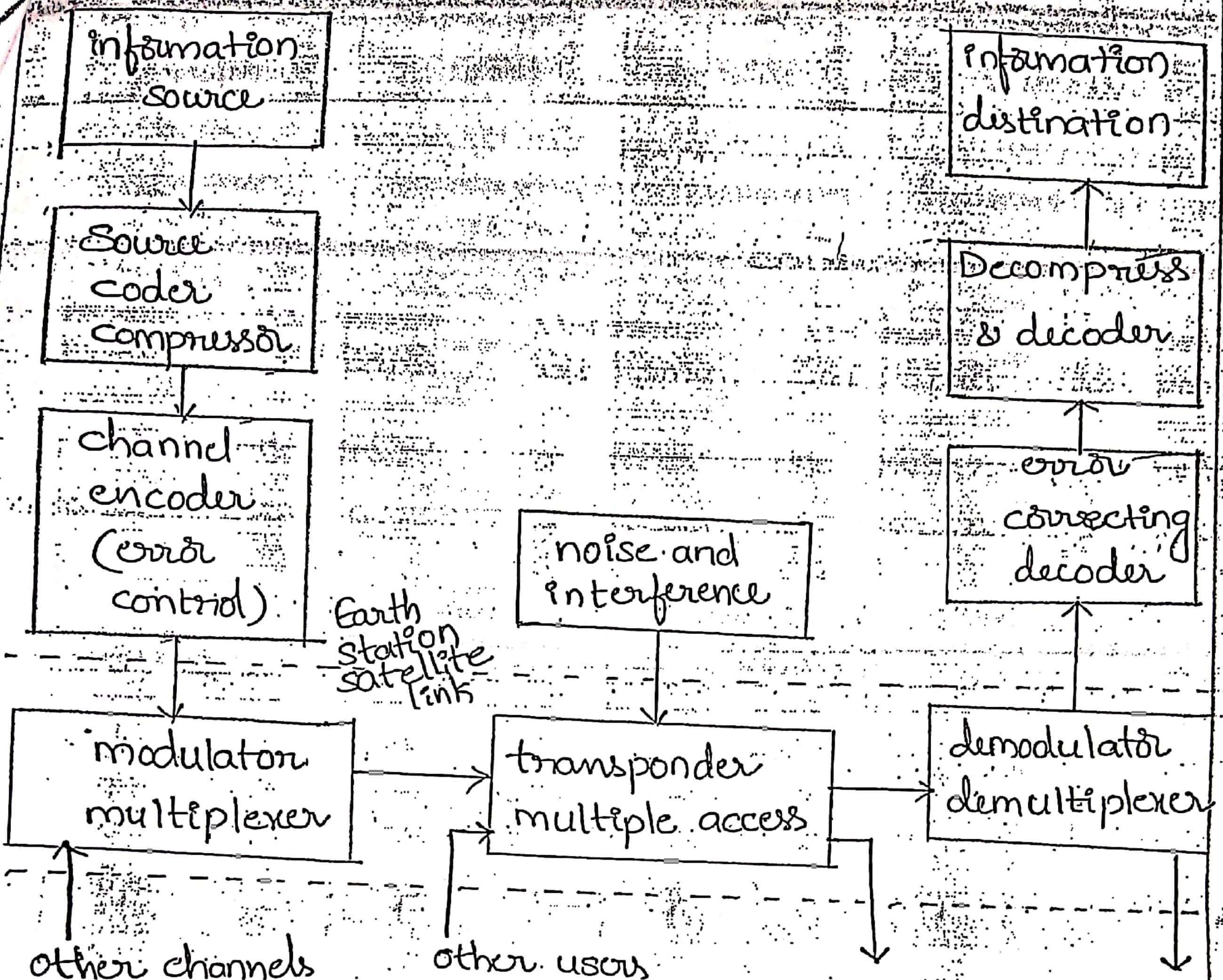


Fig: Prototypical satellite terrestrial link

Above figure represents a generalized communications system including a satellite transmission link. We start with a source of information that can be video, audio or data. The information can be analog or digital. If the transmission or processing is to be digital, the source information is digitized, made as compact as possible without losing any by further source coding, and then compressed to reduce the transmission requirements at the expense of some acceptable loss in quality. Afte every bit of redundancy is squeezed out of the signal, some is then added back selectively in the channel encoder to minimize transmission errors by correcting them using the redundant or parity check bits. The operations of digitizing, source compacting, compression, and channel error correction can be used in various combinations or not at all. Channel encoding ~~encoding~~ is an important area of interaction between the baseband and satellite transmission engineers. Terrestrial transmission often uses simple error detection codes in which a block of digits or a series of blocks is retransmitted on request if an error is detected. The roundtrip time delay on a geostationary satellite link on the order of 0.6s, makes this difficult to effect without serious loss in throughput. But in practice we use only forward error correcting codes that require no retransmission requests. This technique has been made practical and economic in the last decade by the development of high rate convolutional coders and decoders.

The processed digital or analog signal must now modulate a radio frequency signal suitable for transmission to the satellite. Digital baseband signals use phase shift (PSK) or frequency shift keying (FSK) and analog signals use frequency modulation (FM). SSB has been used when the base band signal comprises a large no. of single sideband frequency division multiplexed telephone channels.

The satellite channel is corrupted by thermal, cosmic and device noise, terrestrial and other satellite interference and various linear and nonlinear transmission impairments before undergoing the reverse series of operations to end up at the final destination in whatever form is desired, normally a replica of what was transmitted. It is the function of the demodulators, decoders and decompressors to recover the signal.

If the satellite is to be used for the broadcast of video or audio signals over a wide area to many users, a single transmission to the satellite is repeated and received by multiple receivers. This is a common application of satellites, but there are others in which it is desired to exploit the unique capability of a satellite medium to create an instan network and connectivity between any two points within its view. To exploit this geometric advantage, it is necessary to create some system of multiple access in which many transmitters can use the same transponder simultaneously.

There are three methods of doing this:

FDMA, assigns each transmitter its own carrier frequency. They all transmit simultaneously. Receivers select the desired transmitter by filtering its carrier frequency. This method is extensively used in fixed telephone services to transmit both digital and analog signals.

TDMA is much used today for digital transmission. Each separate transmitter is given its own time slot. The stations transmit sequentially in assigned time slots and all at the same carrier frequency.

CDMA is least used today. In this all the transmitters transmit simultaneously and at the same frequency. Each transmitted signal is modulated by its own pseudo randomly coded bit stream. It is well suited to low data rate systems in crowded spectra.

Frequency allocations for satellite services :-

Allocating frequencies to satellite services is a complicated process which requires international coordination and planning. This is carried out under the auspices of the International Telecommunication Union (ITU).

To facilitate frequency planning, the world is divided into 3 regions.

Region 1 : Europe, Africa (Soviet Union), Mongolia
Region 2 : North & South America, Greenland
Region 3 : Asia (excluding region 1 areas),
Australia, South west Pacific.

Within these regions, frequency bands are allocated to various satellite services, although a given service may be allocated different frequency bands in different regions. Some of the services provided by satellites are

- Fixed Satellite Service (FSS)
- Broadcasting Satellite Service (BSS)
- Mobile Satellite Service

- Navigational Satellite Services
- Meteorological Satellite Services

There are many subdivisions within these broad classifications. For example

the FSS provides links for existing telephone networks as well as for transmitting television signals to cable companies for distribution over cable systems.

Broadcasting Satellite Services are intended mainly for direct broadcast to the home, sometimes referred to as direct broadcast Satellite (DBS) service (in Europe it is known as direct to home (DTH))

Mobile satellite services would include landmobile, maritime mobile, and aeronautical mobile.

Navigational Satellite Services include Global Positioning Systems (GPS) and Satellites intended for the meteorological services often provide a search and rescue service.

Frequency band designations

Frequency range (GHz)	Band designation
0.1-0.3	VHF
0.3-1	UHF
1-2	L
2-4	S
4-8	C
8-12	X
12-18	Ku
18-27	K
27-40	Ka
40-75	V
75-110	W
110-300	mm
300-3000	um

ITU frequency band designations

Band number	Symbols	Frequency ran
4	VLF	3-30KHz
5	LF	30-300KHz
6	MF	300-3000KHz
7	HF	3-30MHz
8	VHF	30-300MHz
9	UHF	300-3000MH
10	SHF	3-30GHz
11	EHF	30-300GH
12		300-3000GH

(lower limit exclusive,
upper limit inclusive)

The Ku band is the one used at present for DBS, and also for certain FSS.

* The C-band is used for FSS and no DBS allowed in this band.
* The VHF band is used for certain mobile and navigational services and for data transfer from weather satellites.
* The L band is used for mobile satellite services and navigation systems.
* For the FSS in the C band the most widely used subrange is approximately 4 to 6GHZ.
* For the DBS in the Ku band, the most widely used range is approximately 12 to 14GHZ.

Future Trends:-

Considering the FSS first, on the global scene the traffic demand is likely to continue to grow, although the advent of transoceanic fibre optic cables may have an impact on the rate of growth. The use of very small aperture terminals for business and rural applications is expected to grow in many parts of the world

Satellites are likely to play a greater role in mobile communications. Mobile satellite systems using non geostationary orbit may begin to emerge towards the second half of the 1990s. The developemen of small hand held terminals which communicate via satellites has been initiated by large service providers such as INMARSAT (International Marine/ Maritime Satellite). Handheld terminals are likely to appear before the year 2000.

The use of direct to home broadcasting is also expected to rise in many parts of the world, although in certain areas satellite broadcast systems may have to compete with other television delivery systems.

For applications such as VSATs or personal mobile terminals, simple inexpensive ground receivers are essential. One possible technical solution is the use of satellites with regenerative repeaters. Such repeaters are more intelligent than the simple repeaters used at present & are equipped with functions such as demodulation and switching. Intelligent satellites together w multiple beam coverage are likely to play an increasing role in the future.

Other areas being investigated include reductio in the coding bit rate of speech signals, which will result in greater bandwidth utilization, the use of as yet un-utilized high radio frequency bands such as 20/30 GHz, to alleviate frequency congestion problems of existing bands, the use of non geostationary orbits for specific applicatio inter satellite links in space to increase space segment capacity and connectivity, advanced antenna concepts, and others.

Applications :-

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

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. Using satellites, to typically reach a distance approximately 10,000kms away, the signal needs to travel almost 72000kms i.e sending data from ground to satellite and from satellite to another location on earth. This causes substantial amount of delay and this delay becomes more prominent for users during voice calls.

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 (ex: researchers on Antartica) or because of the current state of the infrastructure of a country. Here the satellite provides a complete coverage and there is one satellite always present across a horizon.

the earth station. These satellites are exceptiona useful in predicting disasters like hurricanes and monitor the changes in the Earth's vegetation, sea state, ocean color and ice fields.

Radio and TV broadcast :- These dedicated satellites are responsible for making 100s of channels across the globe available for everyone. These are also responsible for broadcasting live matches, news, world wide radio services.

Military satellites :- These satellites are often used for gathering intelligence, as a communication 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.

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 aircrafts rely on GPS as an addition to traditional navigation systems. Many vehicles come with installed GPS receivers.

Global mobile communication :- The basic purpose of satellites for mobile communication is to extend the area of coverage. Cellular phone systems, such as AMPS (Advanced Mobile Phone System), GSM (Global System for Mobile) 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).

Orbital Mechanics :

To achieve a stable orbit around the earth, a spacecraft must first be beyond the bulk of the earth's atmosphere i.e in what is popularly called space. The fundamental Newtonian eqns that describe the motion, of a body are

$$S = ut + \frac{1}{2}at^2 \rightarrow ①$$

$$v^2 = u^2 + 2at \longrightarrow ②$$

$$v = u + at \longrightarrow ③$$

$$P = ma \longrightarrow ④$$

where S = distance travelled from time $t=0$

u = initial velocity of the object at $t=0$

v = final velocity of the object at time t

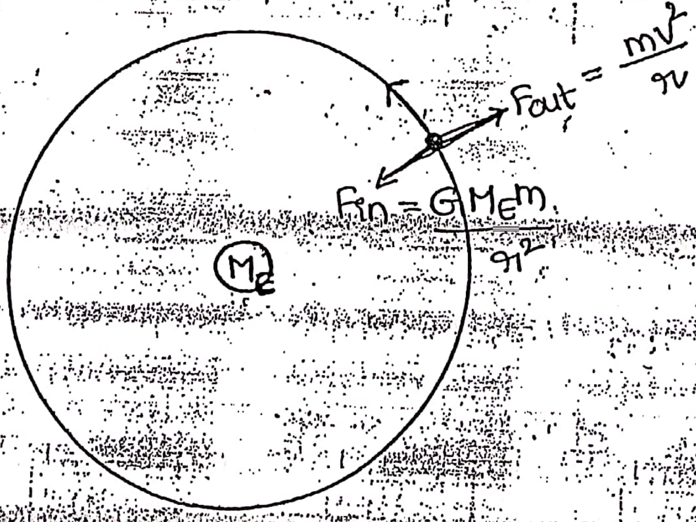
a = acceleration of the object (+ or −)

P = Force acting on the object

m = mass of the object

Of these four eqns it is the last one that helps us understand the motion of a satellite in a stable orbit.

$P = ma$ → states that the force acting on a body is equal to the mass of the body multiplied by the resulting acceleration of the body. thus for a given force, the lighter the mass of the body, the higher thae acceleration will be. When in a stable orbit, there are two main forces acting on a satellite, a centrifugal force due to the kinetic energy of the satellite and a centripetal force due to the gravitational attraction of the planet about which the satellite is orbiting. If these two forces are equal, the satellite will remain in a stable orbit. The below figure shows the two opposite forces acting on a satellite in a stable orbit.



The unit of force is Newton with the notation N. Newton is the force required to accelerati a mass of 1Kg with an acceleration of $1m/s^2$. The underlying units of a Newton are therefore $(kg) \times m/s^2$. The standard acceleration due to gravity at the earths surface is $981\ cm/s^2$. This value decreases with height above the earths surface. The acceleration a due to gravity at a distance r from the center of the earth is $a = \frac{\mu}{r^2}\ km/s^2$

where $\mu = GM_E$ = Keplers constant

$G = 6.672 \times 10^{-11}\ Nm^2/kg^2$ = Universal gravitationa constant

M_E = mass of the earth

$= 5.972 \times 10^{24}\ kg$

$\mu = 3.98 \times 10^{5}\ km^3/s^2$

The centripetal force acting on the satellite, F_{in} is given by $F_{in} = m \times \frac{\mu}{r^2}$

$$= m \times \frac{GM_E}{r^2}$$

The centrifugal acceleration is given by $a = \frac{v^2}{r}$

$$F_{out} = m \times \frac{v^2}{r}$$

If the forces on the satellite are balanced

$F_{in} = F_{out} \Rightarrow m \times \frac{\mu}{r^2} = m \times \frac{v^2}{r}$

Hence the velocity v of a satellite in a circular orbit is given by $v = \sqrt{\frac{\mu}{r}} = \left(\frac{\mu}{r}\right)^{1/2}$

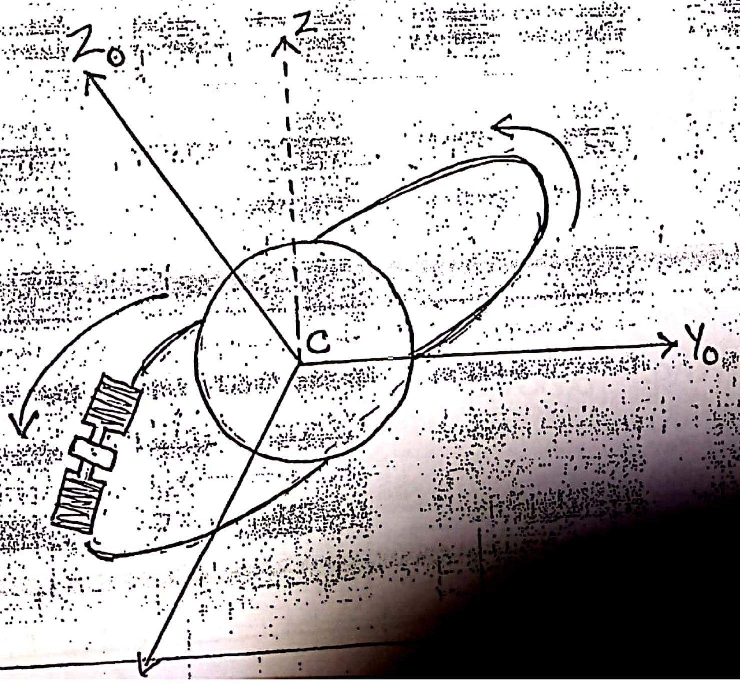
If the orbit is circular, the distance traveled by a satellite in one orbit around a planet is $2\pi r$, where r is the radius of the orbit from the satellite to the center of the planet. The period of the satellites orbit T will be $T = \frac{2\pi r}{v} = \frac{\text{distance travelled}}{\text{velocity}}$

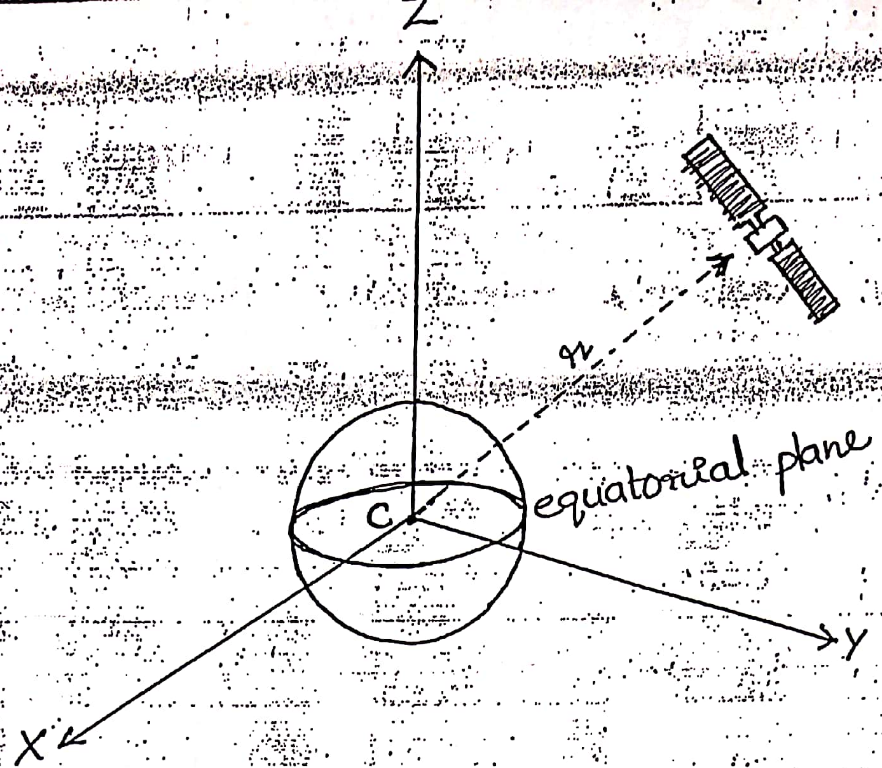
$$= \frac{2\pi r}{\sqrt{\mu/r}} = \frac{2\pi r\sqrt{r}}{\sqrt{\mu}} = \frac{2\pi r^{3/2}}{\mu^{1/2}}$$

A no of coordinate systems and reference planes can be used to describe the orbit of a satellite around a planet. Figure below illustrates one of these using a cartesian coordinate system with the earth at the center and the reference planes coinciding with the equator and the polar axis. This is referred to as a geocentric coordinate system.

$$\frac{d^2\bar{r}}{dt^2} - \frac{\mu}{r^3}\bar{r} = 0 \qquad (\because GM_E = \mu)$$

This is a second order linear differential equation and its solution will involve six undetermined constants called the orbital elements. ($\because$ 2nd order D.E have two solutions involving $r = f(x, y, z)$ i.e $2 \times 3 = 6$ elements) The solution to above eqn is difficult since the 2nd derivative of r involves the second derivative of the unit vector r. To remove this dependence, a different set of coordinates can be chosen to describe the location of the satellite such that the unit vectors in the three axes are constant. This coordinate system uses the plane of the satellites orbit as the reference plane.





With the coordinate system setup as in figure above and with the satellite mass m located at a vector distance r from the center of the earth, the gravitational force F, on the satellite is given by $\bar{F} = \frac{G M_E m \bar{r}}{r^3}$

where M_E = mass of the earth

$G = 6.672 \times 10^{-11}\ Nm^2 / Kg^2$

But force = mass x acceleration and above eqn can be written as $\bar{F} = m \cdot \frac{d^2\bar{r}}{dt^2}$

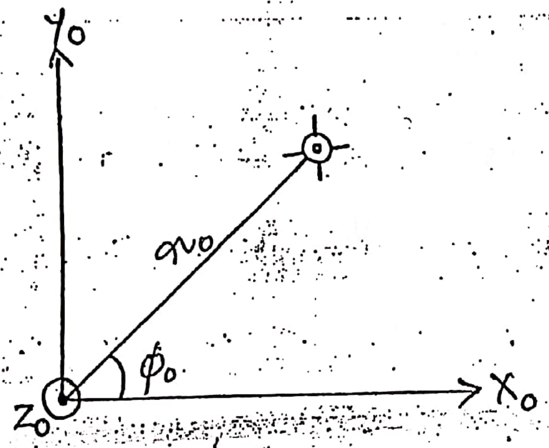
equating above eqns $\frac{G M_E \cancel{m} \bar{r}}{r^3} = \cancel{m} \cdot \frac{d^2\bar{r}}{dt^2}$

$\Rightarrow \frac{d^2\bar{r}}{dt^2} - \frac{G M_E}{r^3}\bar{r} = 0$

Expressing $\frac{d^2 \bar{r}}{dt^2} + \frac{\bar{r}}{r^3} u = 0$ in terms of new coordinate axes x_0, y_0, z_0 gives

$$\hat{x}_0 \left(\frac{d^2 x_0}{dt^2}\right) + \hat{y}_0 \left(\frac{d^2 y_0}{dt^2}\right) + \frac{u\left(x_0 \hat{x}_0 + y_0 \hat{y}_0\right)}{\left(x_0^2 + y_0^2\right)^{3/2}} = 0$$

The above eqn is easier to solve if it is expressed in a polar coordinate system rather than a cartesian coordinate system. The polar coordinate system is shown in below figure



$$\begin{bmatrix} \hat{x}_0 \\ \hat{y}_0 \end{bmatrix} = \begin{bmatrix} \cos\phi_0 & -\sin\phi_0 \\ \sin\phi_0 & \cos\phi_0 \end{bmatrix} \begin{bmatrix} \hat{r}_0 \\ \hat{\phi}_0 \end{bmatrix}$$

With the polar coordinate system shown in above figure and using the transformations

$$x_0 = r_0 \cos\phi_0 \qquad \hat{x}_0 = \hat{r}_0 \cos\phi_0 - \hat{\phi}_0 \sin\phi_0$$

$$y_0 = r_0 \sin\phi_0 \qquad \hat{y}_0 = \hat{\phi}_0 \cos\phi_0 + \hat{r}_0 \sin\phi_0$$

and equating the vector components of r_0 and ϕ_0 inturn in above eqn yields

$$\frac{d^2 r_0}{dt^2} - r_0 \left(\frac{d\phi_0}{dt}\right) = \frac{-u}{r_0^2}$$

and $r_0 \frac{d^2\phi_0}{dt^2} + 2\left(\frac{dr_0}{dt}\right)\left(\frac{d\phi_0}{dt}\right) = 0$

using standard mathematical procedures, we can develop an eqn for the radius of the satellites orbit r_0, namely

$$r_0 = \frac{p}{1+e\cos(\phi_0-\theta_0)}$$

where θ_0 is a constant, e is the eccentricity of an ellipse whose semilatus rectum p is given by $p = \frac{h^2}{\mu}$

h = magnitude of the orbital angular momentum of the satellite

That the eqn of the orbit is an ellipse is Keplers 1st law of planetary motion.

Keplers three laws of planetary motion:-

Johannes Kepler was a German astronomer and scientist who developed his 3 laws of planetar motion by careful observations of the behavior of the planets in the solar system over many years, with the help from the Hungarian astronomer Tycho Brahe.

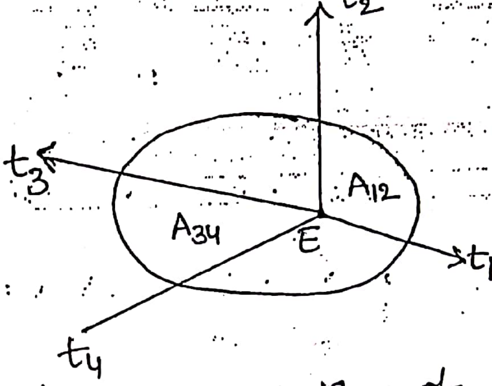
Keplers three laws are

1. The orbit of any smaller body about a larger body is always an ellipse, with the center of mass of the larger body as one of the two foci.

2. The orbit of the smaller body sweeps out equal areas in equal time.

If $t_1 - t_2 = t_3 - t_4$

then $A_{12} = A_{34}$



3. The square of the period of revolution of the smaller body about the larger body equals a constant multiplied by the 3rd power of the semimajor axis of the orbital ellipse. i.e. $T^2 = \frac{4\pi^2 a^3}{\mu}$

where T = orbital period

a = semimajor axis of the orbital ellipse

μ = Keplers constant

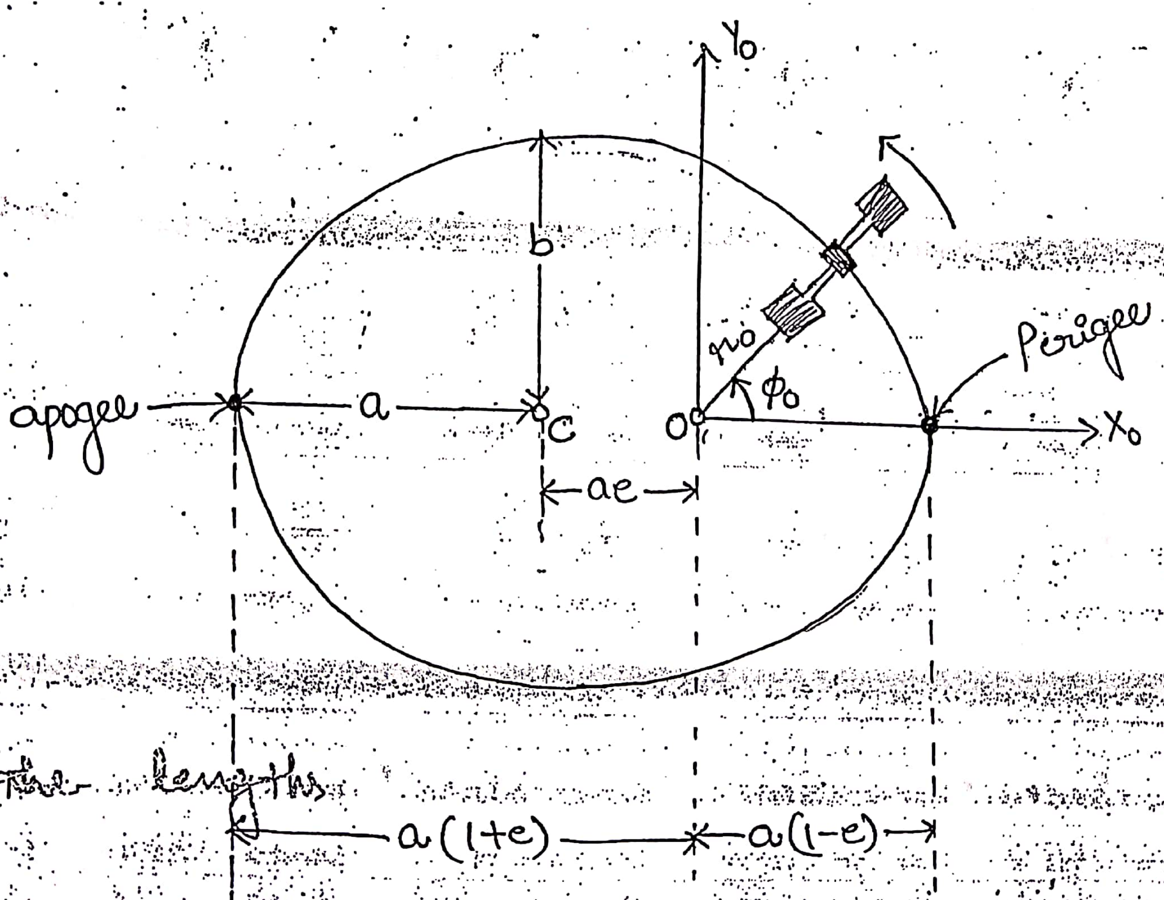
If the orbit is circular, then a becomes distance r.

Describing the orbit of a satellite :-

The quantity θ_0 in eqn $r_0 = \frac{P}{1+e\cos(\phi_0-\theta_0)}$ serves to orient the ellipse wrt the orbital plane axes x_0 and y_0. We know that the orbit is an ellipse, we can always choose x_0 and y_0 so that θ_0 is zero.

$$r_0 = \frac{P}{1+e\cos\phi_0}$$

The path of the satellite in the orbital plane is shown in figure below.



~~The lengths~~

O = center of earth

C = center of the ellipse

The lengths a and b of the semimajor and semiminor axes are given by

$$a = \frac{p}{1-e^2} \qquad b = a(1-e^2)^{1/2}$$

The point in the orbit where the satellite is closest to the earth is called the perigee and the point where the satellite is farthest from the earth is called the apogee. The perigee and apogee are always exactly opposite each other. To make θ_0 equal to zero, we have chosen the x_0 axis so that both the apogee and the perigee lie along it and the x_0 axis is therefore the major axis of the ellipse.

Keplers 3rd law is extremely important in satellite communications. This eqn determines the period of the orbit of any satellite, and it is used in every GPS receiver in the calculation of the positions of GPS satellites. T^2 eqn is also used to find the orbital radius of a GEO satellite, for which the period T must be made exactly equal to the period of one revolution of the earth for the satellite to remain stationary over a point equator.

To be perfectly geostationary, the orbit of a Satellite needs to have three features

(a) it must be exactly circular (i.e eccentricity = 0)

(b) it must be at the correct altitude (i.e have the correct period)

(c) it must be in the plane of the equator (i.e have a zero inclination wrt the equator).

If the inclination of the satellite is not zero and/or if the eccentricity is not zero, but the orbital period is correct, then the satellite will be in a geosynchronous orbit.

Locating the satellite in the orbit : (wrt center of earth)

The equation of the orbit may be written by combining eqns $r_0 = \frac{p}{1+e\cos\phi_0}$ & $a = \frac{p}{1-e^2}$

to obtain $r_0 = \frac{a(1-e^2)}{1+e\cos\phi_0}$

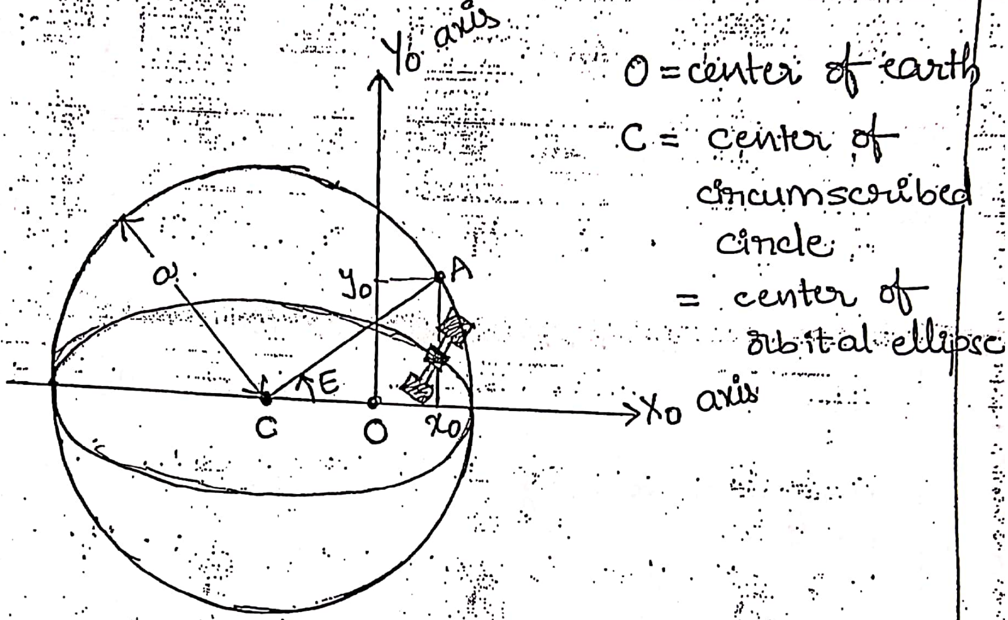
The angle ϕ_0 is measured from the x_0 axis and is called the true anomaly. Since we defined the positive x_0 axis so that it passes through the perigee, ϕ_0 measures the angle from the perigee to the instantaneous position of the satellite. The rectangular coordinates of the satellite are given by $x_0 = r_0 \cos\phi_0$

$y_0 = r_0 \sin\phi_0$

The orbital period T is the time for the satellite to complete a revolution in inertial space, traveling a total of 2π radians. The average angular velocity η is thus $\eta = \frac{2\pi}{T} = \frac{\mu^{1/2}}{a^{3/2}} = \frac{1}{a}\sqrt{\frac{\mu}{a}}$

If the orbit is an ellipse, the instantaneous angular velocity will vary with the position of the satellite around the orbit. If we enclose the elliptical orbit with a circumscribed circle of radius a; then an object going around the circumscribed circle with a constant angular velocity η would complete one revolution in exactly the same period T as the satellite requires to complete one orbital (elliptical) revolution.

Consider the geometry of the circumscribe circle as shown in below fig. Locate the point (indicated as A) where a vertical line drawn through the position of the satellite intersects the circumscribed circle. A line from the center of the ellipse (C) to this point (A) makes an angle E with the x_0 axis; E is called the eccentric anomaly of the satellite.



It is related to the radius r_0 by

$$r_0 = a(1 - e\cos E)$$

Thus $a - r_0 = ae\cos E$

We can also develop an expression that relates eccentric anomaly E to the average angular velocity η, which yields

$$\eta\, dt = (1 - e\cos E)\,dE$$

Let t_p be the time of perigee

We integrati both sides of above eqn

$\eta(t-t_p) = E - e\sin E$

Mean anomaly $M = \eta(t-t_p) = E - e\sin E$ is the arc length (in radians) that the satellite would have traversed since the perigee passage if it were moving on the circum scribed circle at the mean angular velocity η.

If we know the time of perigee t_p, the accentricity e and the length of the semimajor axis a, we now have the necessary equations to determine the coordinates (r_0, ϕ_0) and (x_0, y_0) of the satellite in the orbital plane. The process is as follows.

1. Calculate η using eqn $\eta = \frac{2\pi}{T} = \frac{1}{a}\sqrt{\frac{\mu}{a}}$
2. Calculati M using eqn $M = \eta(t-t_p)$
3. Solve $M = \eta(t-t_p) = E - e\sin E$ for E
4. Find r_0 from E using $a - r_0 = ae\cos E$
5. Solve $r_0 = \frac{a(1-e^2)}{1+e\cos\phi_0}$ for ϕ_0
6. Use $x_0 = r_0\cos\phi_0$, $y_0 = r_0\sin\phi_0$ to calculati x_0 & y_0.

Locating the satellite wrt earth :-

In most cases we need to know where the satellite is from an observation point i.e not at the center of the earth. We will therefore develop the transformations that permit the satellite to be located from a point on the rotating surface of the earth. We will begin with a geocentric equatorial coordinate system as shown in fig below.

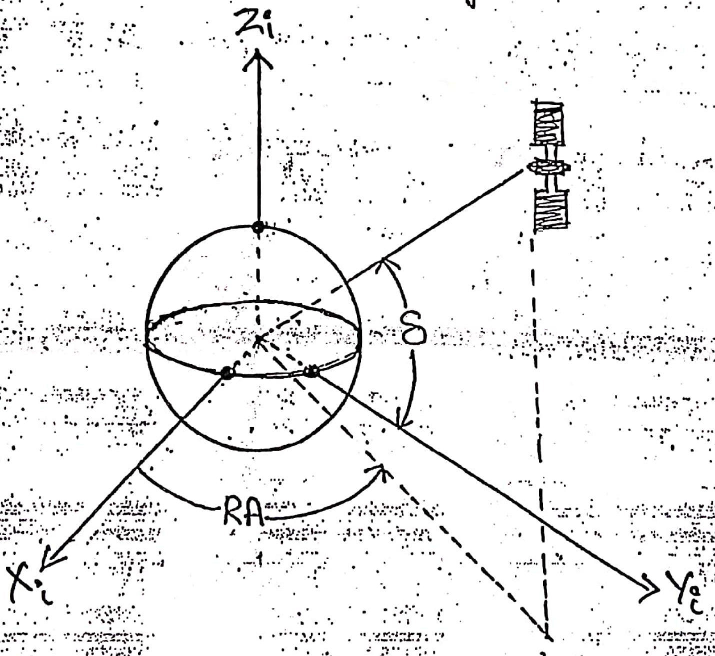
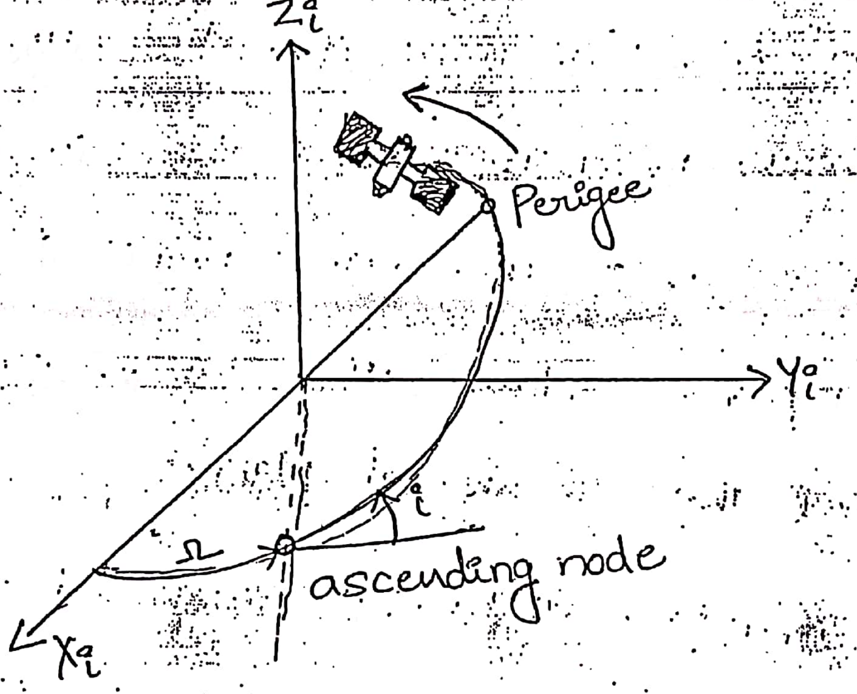


fig: geocentric equatorial system.

The rotational axis of the earth is the Z_i axis, which is through the geographic north pole. The X_i axis from the center of the earth toward a fixed location in space called the first point of aries. This coordinati system moves through space, it translates as the earth moves in its orbit around the sun, but it does not rotate as the earth rotates. The x_i direction is always the same, whatever the earths position around the sun and is in the direction of the first point of Aries. the (x_i, y_i) plane contains the earths equator and is called the equatorial plane.

Angular distance measured eastward in the equatorial plane from the x_i axis is called the right ascension and given the symbol RA. The two points at which the orbit penetrates the equatorial plane are called nodes; the satellite moves upward through the equatorial plane at the ascending node and downward through the equatorial plane at the descending node, given the conventional picture of the earth, with north at the top, which is in the

direction of the positive z axis for the ear
centered coordinate set. The right ascension of the ascending node is called Ω, the angle that the orbital plane makes with the equatorial plane is called the inclination i.



The variables Ω and i together locate the orbital plane wrt the equatorial plane. To locate the orbital coordinate system wrt the equatorial coordinate system we need ω, the argument of perigee west. This is the angle measured along the orbit from the ascending node oft to the perigee.

Standard time for space operations & most other scientific and engineering purposes is universal time (UT) also known as Zulu time (Z). This is essentially the mean solar time at the Greenwich observatory near London, England. Universal time is measured in hours, minutes and seconds. Astronomers employ a second dating system involving Julian days and Julian dates. Julian days starts at noon UT in a counting system. To find exact position of an orbiting satellite at a given instant in time requires knowledge of the orbital elements.

Orbital elements :-

To specify the absolute coordinates of a satellite at time t, we need to know 6 quantities. These quantities are called the orbital elements. Morethan six quantities can be used to describe a unique path and there is some arbitariness in exactly which six quantities are used. We have chosen to adopt a set that is commonly used in satellite communications: eccentricity (e), semi major axis (a), time of perigee (tp), right

Unit-2

16-442

Satellite Subsystem

Communication satellites are very complex, extremely expensive to purchase and to launch. This cost is increased by the need to dedicate an earth station to the monitoring and control of the satellite. The revenue to pay these costs is obtained by selling the communication capacity of the satellite to users. communication satellites are usually designed to have a typical operating life time of 10-15 years. In order to support the comm. systems, the satellite must provide a stable platform on which to mount the antennas, be capable of station keeping, provide the required electrical power to the comm. systems & also provide a controlled temperature environment for comm. electronics. comm. satellites for low earth orbit are in most cases quite similar to small GEO satellites and have similar requirements.

Attitude and Orbit control systems (AOCS):

This subsystem consists of rocket motors that are used to move the satellite back to the correct orbit when external forces cause it to drift off station and gas gets or inertial devices that control the attitude of the satellite. The attitude and orbit of a satellite must be controlled so that the satellite's antennas point toward the earth and so that the user knows where in the sky to look for the satellite

There are several forces acting on an orbiting satellite that tend to change its attitude and orbit. The most important are the gravitational fields of the sun and the moon, irregularities in earth's gravitational field, solar pressure from the sun and variations in the earth's magnetic field.

Solar pressure acting on a satellite's solar sails and antennas and the earth's magnetic field generating eddy currents in the satellite's metallic structure as it travels through the magnetic field, tend to cause rotation of the satellite body. Careful design of the structure can minimize these effects, but the orbital period of the satellite makes many of the effects cyclic, which can cause nutation of the satellite. The attitude control system must damp out nutation and counter any rotational torque or movement.

Attitude control system :

There are two ways to make a satellite stable in orbit, when it is weightless. The body of the satellite can be rotated, to create a gyroscopic force that provides stability of the spin axis and keeps it pointing in the same direction. Such satellites are known as spinners.

Hughes 376 (Boeing 376) → spin stabilized satellite

Hughes 701 (Boeing 701 Series) → three axis stabilized satellite

The satellite can be stabilized by one or more momentum wheels. This is called a three axis stabilized satellite. The spinner design of satellite is typified by many satellites built by Hughes Aircraft Corporation for domestic satellite communication systems. This satellite consists of a cylindrical drum covered in solar cells that contains the power systems and the rocket motors. The comm. system is mounted at the top of the drum and is driven by an electric motor in the opposite direction to the rotation of the satellite body to keep the antennas pointing toward the earth. Such satellites are called despun.

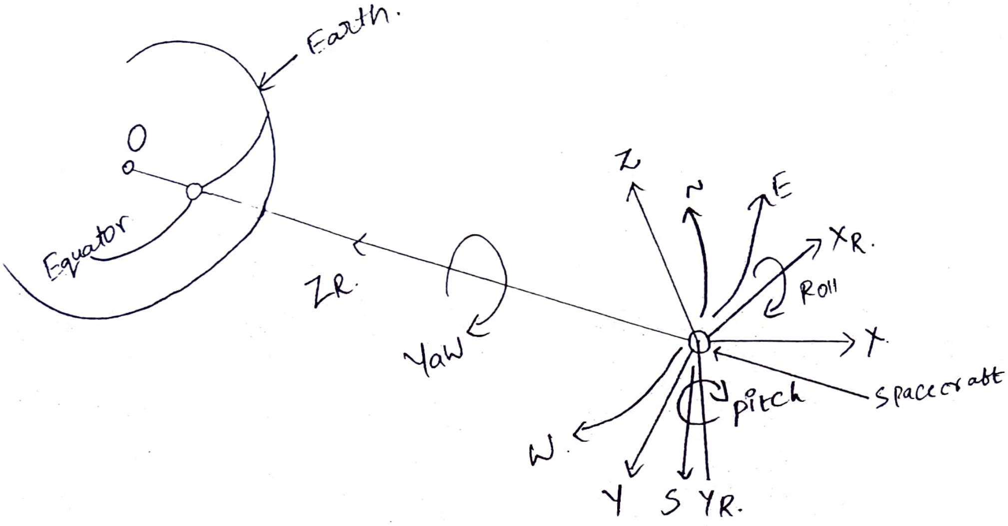
The satellite is spun by operating small radial gas jets mounted on the perephery of the drum, at an appropriate point in the launch phase. The despin system is then brought into operation so that the main TTC&M antennas point toward the earth. A variety of liquid propulsion mixes have been used for the gas jets, the most common being a variant of hydrazine (N_2H_4), which is easily liquefied under pressure, but readily decomposes when passed over a catalyst. Increased power can be obtained from the hydrazine gas jets by electrically heating the catalyst and the gas. The most common bipropellents used for thruster operations are mono methyl hydrazine and nitrogen tetroxide. The bipropellants are hypogolic: i.e they ignite spontaneously on contact and so do not need either a catalyst or a heater. By adjusting the flow of bipropellants, pulses of thrust can be generated at the correct time and in the correct direction.

There are two types of rocket motors used on satellites. The traditional bipropellant thruster described above, and arc jets or ion thrusters. The fuel that is stored on a GEO satellite is used for two purposes: to fire the apogee kick motor that injects the satellite into its final orbit, and to maintain the satellite in that orbit over its lifetime. Arc jets or ion thrusters are mainly used for north-south station keeping.

In a three axis stabilized satellite, one pair of gas jets is needed for each axis to provide for rotation in both directions of pitch, roll, yaw. When motion is required along a given axis, the appropriate gas jet is operated for a specified period of time to achieve the desired velocity. The opposing gas jet must be operated for the same length of time to achieve ~~the~~ stop the motion when the satellite reaches its new position. Fuel is saved if the velocity of the satellite is kept small, but progress toward the destination is slow.

Let us define a set of reference cartesian axes (X_R, Y_R, Z_R) with the satellite at the origin. The Z_R axis is directed toward the center of the earth and is in the plane of the satellite orbit. The X_R axis is tangent

, the orbital plane and lies in the orbital plane. the Y_R axis is ⊥r to the orbital plane.



Rotation about the $X_R, Y_R \cdot Z_R$ axes is defined as roll about the X_R axis, pitch about the Y_R axis and yaw about the $\dot{Z}_R$ axis, in exactly the same way as for an aircraft or ship traveling in the x direction. The satellite must be stabilized wrt the ref axes to maintain accurate pointing of its antenna beams. The axes X_R, Y_R & Z_R are defined wrt the location of the satellite, a second set of cartesian axes, X, Y, Z as shown in fig below, define the orientation of the satellite. changes in a satellites attitude cause the angles θ, ϕ & ψ to vary as the X, Y & Z axes move relative to the fixed reference axes X_R, Y_R & Z_R. The z axis is usually directed toward a ref point on the earth, called the z-axis intercept. The location of the z axis intercept defines the pointing of the satellite antennas.

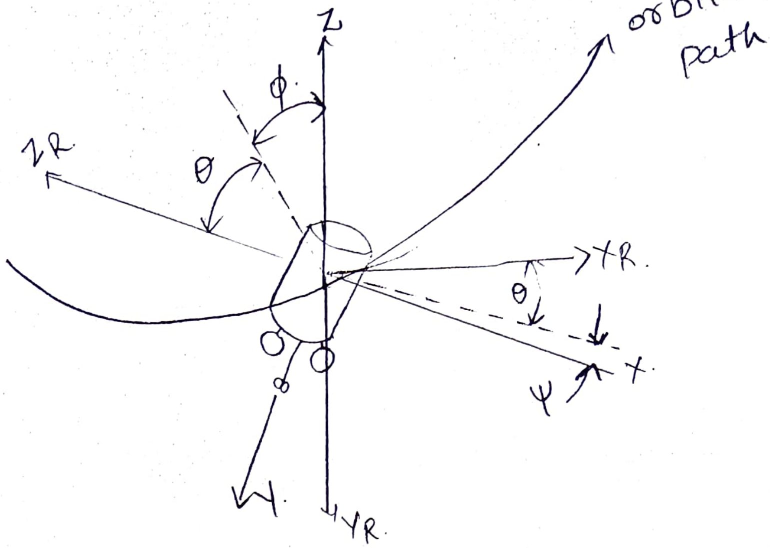


Figure below illustrates how an infrared sensor on the spinning body of a satellite can be used to control pointing toward the earth. the control system will be more complex for a three axis stabilized satellite and may employ an onboard computer to process the sensor data & command the gas gets & momentum wheels.

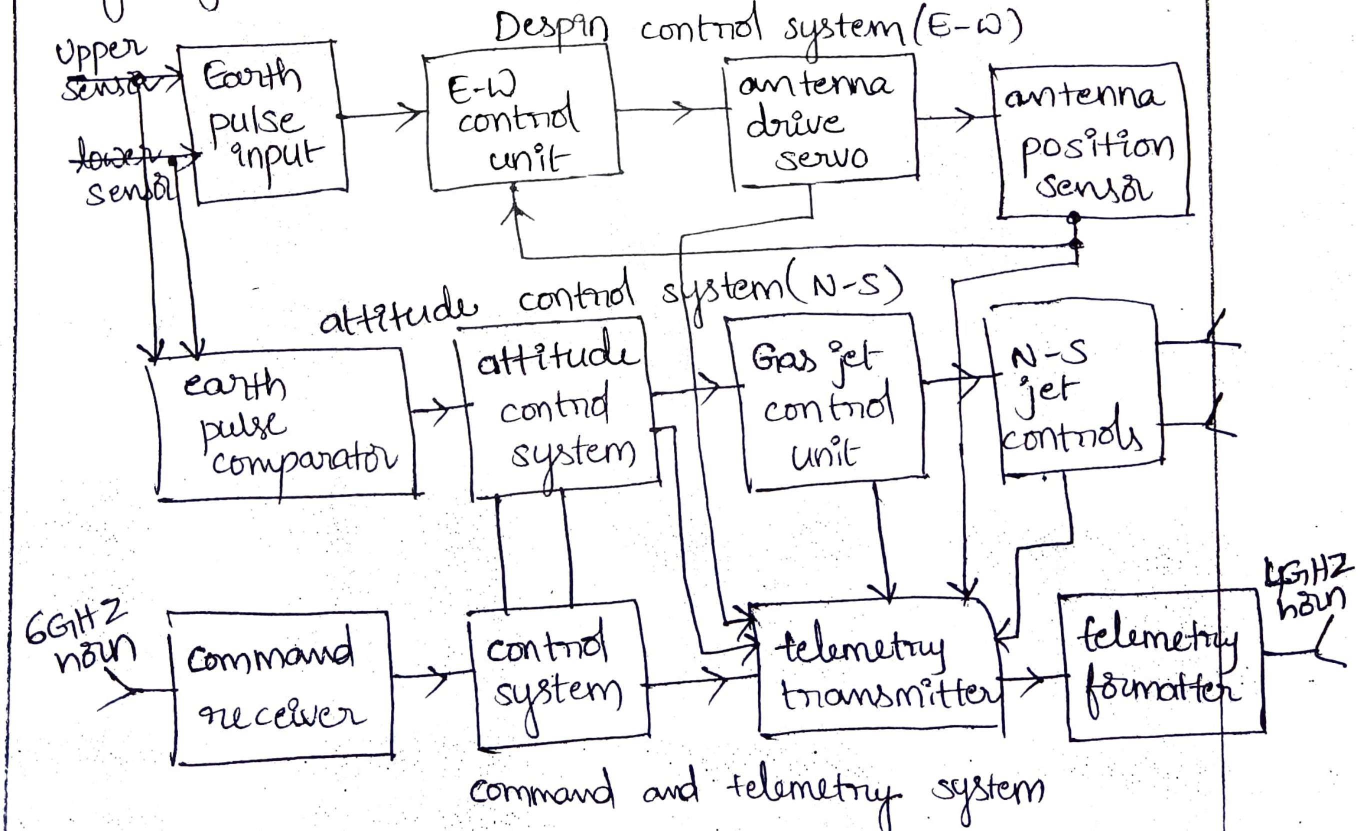


fig: Typical onboard control system for a spinner satellite.

Telemetry, Tracking, Command and monitoring (TTC&M)

The TTC&M system is essential to the successful operation of a comm. satellite. It is the function of it is, to control the orbit and attitude of the satellite, monitor the status of all sensors and subsystems on the satellite and switch on or off sections of the comm. system. The TTC&M earth station may be owned and operated by the satellite owner, or it may be owned by a 3rd party & provide TTC&M services under contract.

Telemetry and monitoring system

The monitoring system collects data from many sensors within the satellite and sends these data to the controlling earth station. There may be several hundred sensors located on the satellite to monitor pressure in the fuel tanks, voltage & current in the power conditioning unit, current drawn by each subsystem and critical voltages & currents in the comm. electronics. The temperature of many of the subsystems is important and must be kept within predetermined limits, so many temperature sensors are fitted. The sensor data, the status of each subsystem, and the positions of switches in comm.

systems are reported back to the earth by the telemetry system.

Telemetry data are usually digitized and transmit as PSK of a low power telemetry carrier using time division techniques. A low data rate is normally used to allow the receiver at the earth station to have a narrow bandwidth and thus maintain a high carrier to noise ratio. At the controlling earth station a computer can be used to monitor, store and decode the telemetry data so that the status of any system or sensor on the satellite can be determined immediately by the controller on the earth. Alarms can also be sounded if any vital parameter goes outside allowable limits.

Tracking :

A no.of techniques can be used to determine the current orbit of a satellite. velocity and acceleration sensors on the satellite can be used to establish the change in orbit from the last known position, by integration of data. The earth station controlling the satellite can observe the doppler shift of the telemetry carrier or beacon transmitter carrier to determine the rate at which range is changing. Together with accurate angular measurements from the earth station

Smith

antenna, range is used to determine the orbital elements. Active determination of range can be achieved by transmitting a pulse or sequence of pulses, to the satellite and observing the time delay before the pulse is received again. The propagation delay in the satellite transponder must be accurately known, and morethan one earth station may make range measurements. With precision equipment at the earth stations, the position of the satellite can be determined within 10m.

Ranging tones are also used for range measurement. A carrier generated on board the satellite is modulated with a series of sine waves at increasing frequency, usually harmonically related. The phase of the sine wave modulation components is compared at an earth station, and the no.of wavelengths of each frequency is calculated. Ambiguities in the numbers are resolved by reference to lower frequencies, and prior knowledge of the approx range of the satellite. If sufficiently high frequencies are used, perhaps even the carrier frequency, range can be measured to mm accuracy. The technique is similar to that used in the terrestrial telurometer and in aircraft radar altimeters.

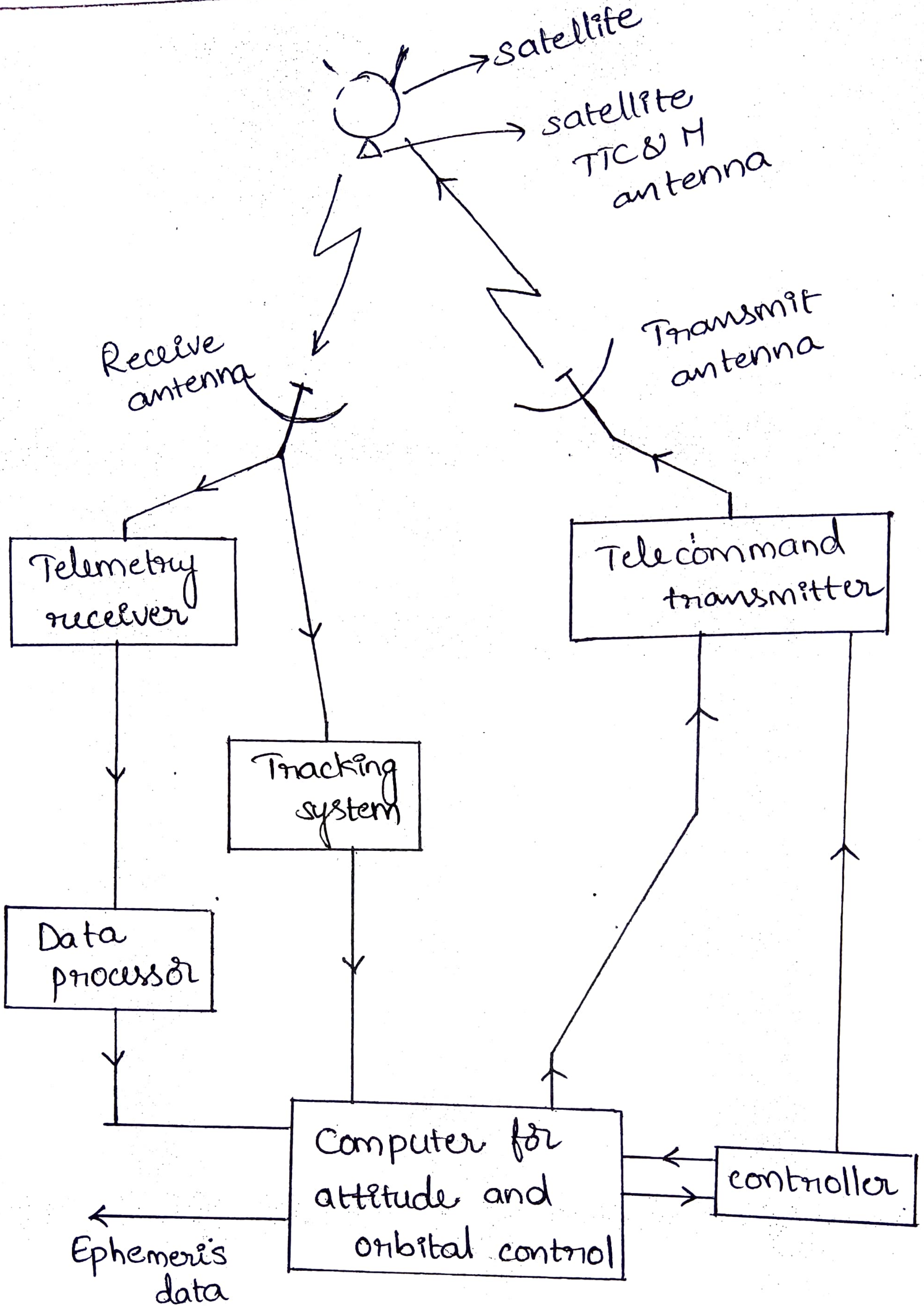


fig: Typical tracking, telemetry, command and monitoring system.

Command :-

A secure and effective command structure is vital to the successful launch and operation of any comm. satellite. The command structure must posses safeguards against unauthorized attempts to make changes to the satellite's operation. Encryption of commands and responses is used to provide security in the command system. the control code is converted into a command word, which is sent in a TDM frame to the satellite. After checking for validity in the satellite, the word is sent back to the control station via the telemetry link where it is checked again in the computer. If it is found to have been received correctly, an execute instruction will be sent to the satellite so that the command is executed. The entire process may take 5 or 10s, but minimizes the risk of erroneous commands causing a satellite malfunction.

During the launch phase and injection into geostationary orbit, the main TTC&M system may be inoperable because the satellite does not have the correct attitude or has not extended its solar sails. A back up system is used at this time, which controls only the most important sections of the satellite. The back up system provides control of the apogee kick motor, the attitude control system and orbit control thruste the solar sail deployment mechanism and the power conditioning unit. With these controls, the satellite can be injected into geostationary orbit, turned to face the earth and switched to full electrical power so that handover to the main TTC&M system is possible. In the event of failure of the main TTC&M system, the backup system can be used to keep the satellite on station. It is also used to eject the satellite from geostationary orbit and to switch off all transmitters when the satellite eventually reaches the end of its useful life.

Power systems:

All comm. satellites obtain their electrical power from solar cells. Some deep space planetary research satellites have used thermonuclear generators to supply electrical power, but because of the danger to people on the earth if the launch should fail & the nuclear fuel be spread over an inhabited area comm. satellites have not used nuclear generators.

The sun is a powerful source of energy. In the total vacuum of outer space, at geostationary altitude, the radiation falling on a satellite has an intensity of $1.39 KW/m^2$. Solar cells do not convert all this incident energy into electrical power. Their efficiency is typically 20-25% at beginning of life (BOL) but falls with time because of aging of the cells and etching of the surface by micrometeor impacts. Since sufficient power must be available at the end of life (EOL) of the satellite. about 15% extra area of solar cells is usually provided as an allowance for aging.

A spin stabilized satellite usually has a cylindrical body covered in solar cells. Because the solar cells are on a cylindrical surface, half of the cells are not illuminated at all, and at the edges of the illumenated half, the low angle of incidence results in little electrical power being generated. The output from the solar cells is slightly higher than would be obtained with normal incidence on a flat panel equal ~~to~~ in area to the projected area of the cylinder i.e its width times its height. Early satellites were of small dimensions and had relatively small areas of solar cells. Recently, comm. satellites for direct broadcast operation generate upto 6KW from solar power.

A three axis stabilized satellite can make better use of its solar cell area, since the cells can be arranged on flat panels that can be rotated to maintai normal incidence of the sunlight. A primary advantage is that by unfurling a folded solar array, power in excess of 10KW can be generated with large arrays. Solar sails must be rotated by an electric motor once per 24H to keep the cells in full sunlight. This causes the cells to heat up, typically to 50° to 80°C, which causes a drop in output voltage. In the spinner design, the cells cool down when in shadow and run at 20° to 30°C with somewhat higher efficiency.

The satellite must carry batteries to power the subsystems during launch and during eclipses. Eclipses occur twice per year, around the spring and fall equinoxes. The longest duration of eclipse is 70 min, occuring around March 21 and september 21 each year. Batteries are usually of the nickel-hydrogen type which do not gas when charging and have good reliability and long life and can be safely discharged to 70% of their capacity. A power conditioning unit controls the charging current and dumps excess current from the solar cells into heaters or load resistors on the cold side of the satellite. Sensors on the batteries, power regulator, and solar cells monitor temperature, voltage and current and supply these data to both the onboard control system and the controlling earth station via the telemetry downlink. Typical battery voltages are 20-50V with capacities of 20-100 ampere-hours.

Communications subsystems

The comm. sub system is the major component of a comm. satellite & the remainder of the satellite is there solely to support it. It is usually composed of one or more antennas & a set of receivers & transmitter. Since it is the comm. system that earns the revenue for the system operator, comm. satellites are designed to provide the largest traffic capacity possible. Successive satellites have become larger, heavier, and more costly, but the rate at which traffic capacity has increased been much greater, resulting in a lower cost per telephone circuit. The satellite transponders have limited output power and the earth stations are atleast 36,000km away from a GEO satellite, so that received power p is very small and rarely exceeds 10^{-10}W. For the system to perform satisfactorily, the signal power must exceed the noise power generated in the receiver by between 5 & 25db, depending on the B.W of the Txed signal & the modulation scheme used. With low power transmitters, narrow receiver bandwidths have to be used to maintain the required S/N ratio.

Early comm. satellites were fitted with transponders of 250 or 500MHZ bandwidth, but had low gain antennas and transmitters of 1 or 2W output power. The earth station receiver could not acheive an adequate S/N ration when the full bandwidth was used with the result that the system was power limited. Later generations of comm. satellites have transponders with greatly increased output power upto 200W for DBS-TV satellites & have steadily improved in bandwidth utilization efficiency. The total channel capacity of a satellite can be increased only if the B.W can be increased or reused. The trend in high capacity satellites has been to reuse the available bands by employing several directional beams at the same frequency (spatial freq reuse) and orthogonal polarizations at the same freq (polarization freq reuse).

The designer of a satellite comm. system is not free to select any frequency and bandwidth he or she chooses International agreements restrict the frequencies that may be used for particular services and the regulations are administered by the appropriate agency in each country the Federal comm. Commission (FCC) in U.S for example The bands currently used for the majority of services are 6/4GHZ & 14/11 GHZ with 30/20GHZ coming into service.

The standard spacing between GEO satellites was originally set at 3°, the spacing has been reduced to 2°. the move to 2° spacing opened up extra slots for new satellites

Transponders :—

Signals transmitted by an earth station are received at the satellite by either a zone beam or a spot beam antenna. Zone beams can receive from transmitters any where with in the coverage zone, where as spot beams have limited coverage. The fig shows a simplified block diagram of a satellite comm. subsystem for the 6/4 GHZ band. the 500MHZ B.W is divided up into channels, often 36MHZ wide, which are each handled by a separate transponder. A transponder consists of a BPF to sel the particular channel's band of frequencies, a downconv to change the frequency from 6GHZ at the input to 4GHZ at the output, and an output amplifier. The comm. system has many transponders, some of which may be spares, typically 12-44 active transponders are carried by a high capacity satellite. The transponders are supplied with signals from one or more receive antennas and send their outputs to a switch matrix that directs each transponder band of frequencies to the appropriate antenna or antenna beam. The switch setting can be controlled from the earth to allow reallocation of the transponders between the downlink beams as traffic patterns change.

Many domestic satellites operating in the 6/4 GHZ band carry 24 active transponders. The center freq of the transponders are spaced 40MHZ apart, to allow guard bands for the 36MHZ filter skirts. With a total of 500MHZ available, a single polarization satellite can accomodate 12 transponders across the band. When freq reuse by orthogonal polarizations is adopted, 24 transponders can be accommodated in the same 500MHZ bandwidth. The reuse is achieved through microwave switch interconnections between subbeams. The only way to achieve this level of beam/path interconnections is via on board processing (OBP).

When morethan one signal shares a transponder using FDMA, the power amplifier must be run below its maximum output power to maintain linearity and reduce intermodulation products. The degree to which the transmitter output power is reduced below it peak output is known as output backoff, in FDMA systems, 2 to 7db of output backoff is typically used. TDMA can theoretically be used to increase the output power of transponders by limiting the transponder to a single access. Most TDMA systems are hybrid FDMA - TDMA schemes known as multi frequency TDMA (MF-TDMA), in which several TDMA signals share the transponder bandwidth using FDMA.

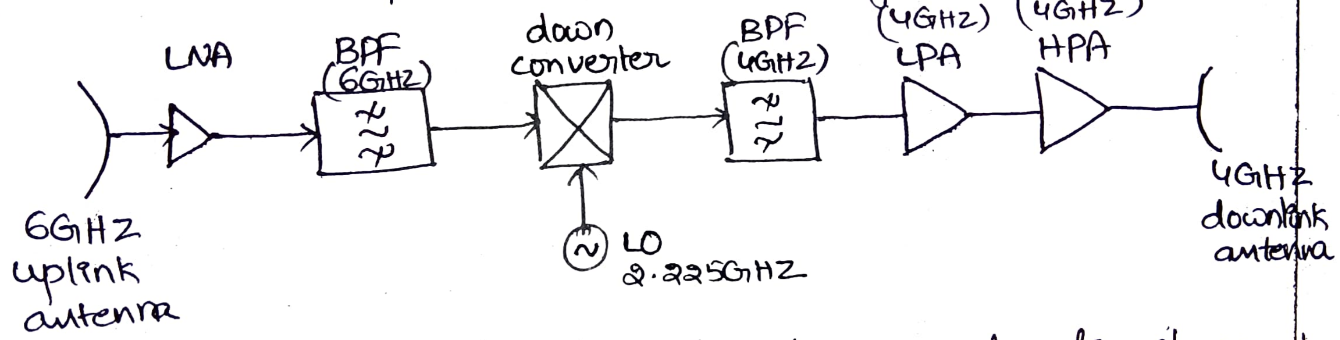
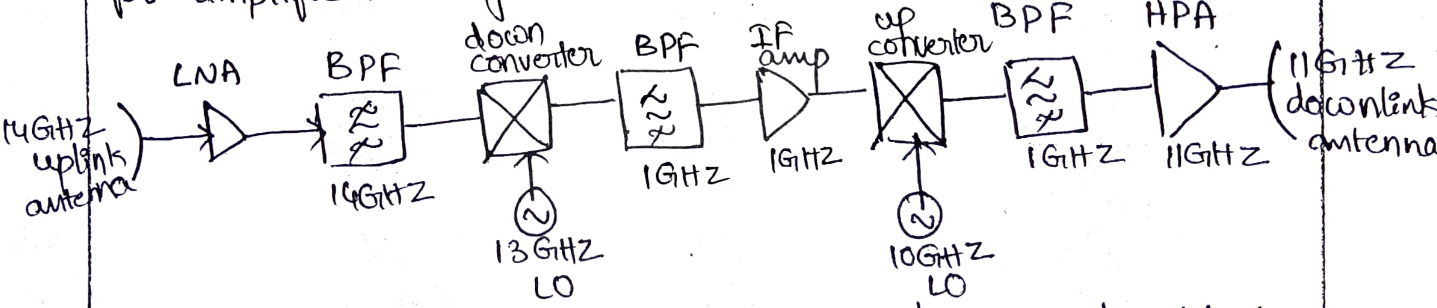


fig: Simplified single conversion transponder for 6/4 GHz band

Fig shows a typical single conversion bent pipe transponder of the type used on many satellites for the 6/4 GHZ band. The output power amplifier is usually a solid state power amplifier (SSPA), unless a very high output power (>50W) is required, when a TWT amplifier would be used. The LO is at 2225MHZ to provide the appropriate shift in freq from 6GHZ uplink freq to the 4GHZ downlink freq & the BPF after the mixer removes unwanted freq resulting from down conversion. Redundancy is provided for the HPA in each transponder by includin a spare TWT or SSPA that can be switched into circ if the primary power amplifier fails. Transponders can also be arranged so that there are spare transponders available in the event of a total failure. The arrangement is known as M for N redundancy. (For example 16 for 10 redundancy, 10 → active 6 → spare)

Transponders for use in the 14/11 GHZ bands normally employ a double freq conversion scheme. It is easier to make filters, amplifiers and equalizers at an IF such as 1100MHZ than at 14 or 11GHZ. So the incoming 14GHZ carrier is translated to an IF of around 1GHZ. The amplification & filtering are performed at 1GHZ and a relatively high level carrier is translated back to 11GHZ for amplification by the HPA.

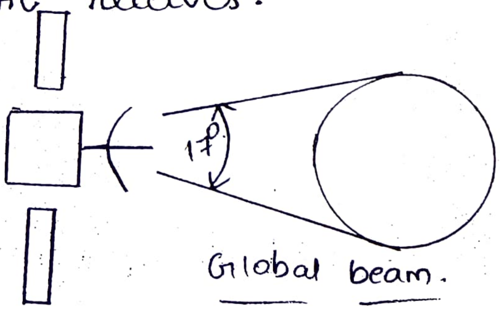


It is possible to conserve uplink bandwidth by using different modulation techniques on the uplink and downlink and by providing a baseband processor on the satellite. A high level modulation such as 16-QAM with four bits per symbol can be used on the link between the satellite and a large earth station to improve bandwidth efficiency. This approach has been adopted in the Astrolink and spaceway 30/20 GHZ satellites.

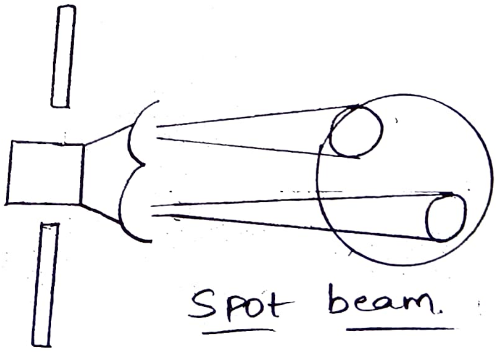
Satellite antennas

Four main types of antennas are used on satellites. These are
1. Wire antennas : monopoles and dipoles
2. Horn antennas
3. Reflector antennas
4. Array antennas

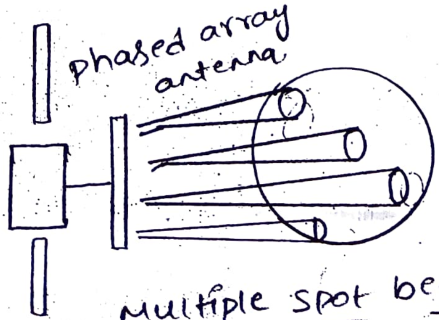
Wire antennas are used primarily at VHF & UHF to provide comm. for the TTC&M systems. An antenna pattern is a plot of the field strength in the far field of the antenna when the antenna is driven by a transmitter. It is usually measured in db. The gain of an antenna is a measure of the antenna's capability to direct energy in one direction, rather than all around. Reciprocity means that an antenna has the same gain and pattern at any given freq whether it transmits or receives.



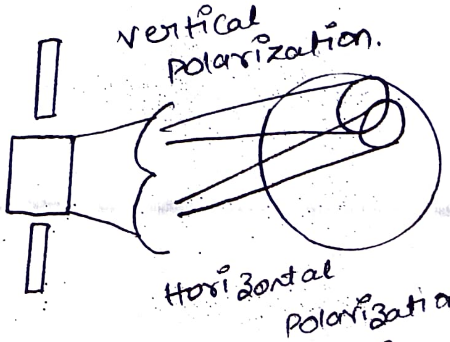
Global beam.



Spot beam.



Multiple spot beams and scanning beams.



orthogonally polarized beams

Fig. shows typical satellite antenna coverage zones. The pattern is frequently specified by its 3db beamwidth, the angle between the directions in which the radiated or received field falls to half the power in the direction of maximum field strength.

Horn antennas are used at microwave frequencies when relatively wide beams are required, as for global coverage. A horn is a flared section of waveguide that provides an aperture several wavelengths wide and a good match between the waveguide impedance and free space. Horns are also used as feeds for reflectors, either singly or in clusters. Horns and reflectors are examples of aperture antennas that launch a wave into free space from a waveguide. It is difficult to obtain gains much greater than 23db or beamwidths narrower than about 10° with horn antennas. For higher gains or narrow beamwidths a reflector antenna or array must be used.

Reflector antennas are usually illuminated by one or more horns and provide a larger aperture than can be achieved with a horn alone. Phased array antennas are also used on satellites to create multiple beams from a single aperture. Some basic relationships in aperture antennas can be used to determine the approximate size of a satellite antenna for a particular application, as well as the antenna gain.

n aperture antenna has a gain G given by

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

where A = area of the antenna aperture in meters

λ = operating wavelength

η_A = aperture efficiency

η_A = 55-68% for reflector antennas

= 65-80% for horn antennas

If the aperture is circular then $G = \eta_A \left(\frac{\pi D}{\lambda}\right)^2$

D = diameter of the circular aperture in meters

The beam width of antenna

$$\theta_{3db} \simeq 75 \frac{\lambda}{D} \text{ degrees}$$

Satellite antennas in practice :-

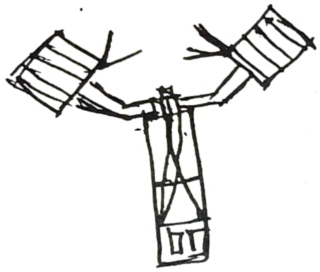
The antennas of a comm. satellite are often
a lemiting element in the complete system. In an ideal
satellite, there would be one antenna beam for each earth
station, completely isolated from all other beams, for
transmit and receive. However, if two earth stations
are 300km apart on the earth's surface and the
satellite is in geostationary orbit, their angular separation
at the satellite is 0.5°. A phased array feed is used
to create many 0.5° beams which can be clustered to
serve the coverage zone of the satellite

To provide a separate beam for each earth station would also require one antenna feed per earth station if a multiple feed antenna with a single reflector were used. A compromise between one beam per station and one beam for all stations has been used in many satellite by using zone coverage beams and orthogonal polarizations with in the same beam to provide more channels per satellite.

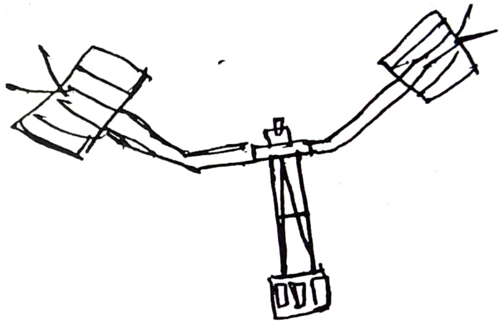
In 2000 the geostationary orbit had domestic satellites spaced every 2°, operating at 6/4 GHz and 14/11 GHz from longitude 68°w to 140°w. This encompasses all orbital locations that can be simultaneously viewed by earth stations in the US & canada, and each operator has been given a limited no.of orbital slots in which to place a satellite. As a result, there is a great deal of pressure on the operating companies to obtain the max no.of channels per satellite in order to give the operator the greatest possible revenue earning capacity. This has encouraged the development of frequency reuse antennas by means of orthogonal polarizations and multiple beams, the combination of 6/4 and 14/11 GHz comm. system on one satellite, and the use of multilevel digital modulation & TDMA to increase capacity. The requirements of narrow antenna beams with high gain over a small coverage zone leads to large antenna structures on the satellite

Frequently, the antennas in their operating configuration are too large to fit within the shroud dimensions of the launch vehicle, and must be folded down during the launch phase. Once in orbit, the antennas then can be deployed. In many larger satellites, the antennas use offset paraboloidal reflectors with clusters of feeds to provide carefully controlled beam shapes the feeds mount on the body of the satellite, close to the comm. subsystem and the reflector is mounted on a hinged arm.

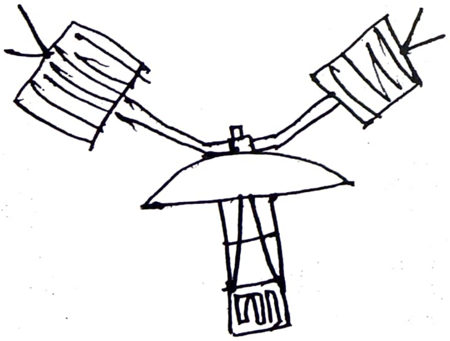
1. After separation



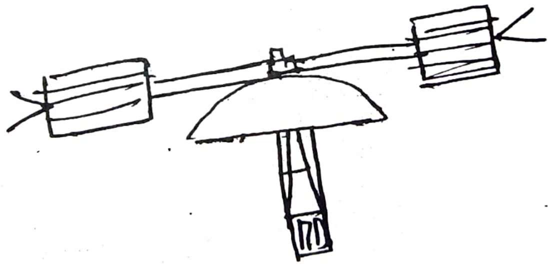
2. Solar array booms extended



3. Solar array panels extended.



4. 30-ft reflector deploys



5. fully deployed configuration.

Fig shows deployment seq used for the 30ft antenna carried by ATS-6: the antenna was built as a series of petals that folded over each other to make a compact unit during launch, which then unfurled in orbit. The solar sails folded down over the antenna and were deployed first. springs or pyrotechnic devices can be used to provide the energy for deployment of antennas or solar sails, with a locking device to ensure correct positioning after deployment.

Equipment reliability and space qualification

Comm. satellites built in 1990s have operational lifetimes of upto 15 years. Once a satellite is in geo-stationary orbit, there is little possibility of repairing components that fail or adding more fuel for station keeping. The components that make up the satellite must therefore have very high reliability in the hostile environment of outer space, and a strategy must be devised that allows some components to fail without causing the entire comm. capacity of the satellite to be lost. Two separate approaches are used:

- Space qualification of every part of the satellite to ensure that it has a long life expectancy in orbit
- redundancy of the most critical components to provide continued operation when one component fails.

Space qualification

The 1st stage in ensuring high reliability in a satellite is by selection and screeing of every component used. Past operational and test experience of components indicates which components can be expected to have good reliability. Only components that have been shown to have high reliability under outer space conditions will be selected. Each component is then tested individually (or as a subsystem) to ensure that it meets its specification. This process is known as quality control or quality assurance. Once individual components and sub-systems have been space qualified, the complete satellite must be tested as a system to ensure that its many systems are reliable.

When a satellite is designed, three prototype models are often built and tested.
1. mechanical model
2. Thermal model
3. electrical model

The mechanical model contains all the structural and mechanical parts that will be included in the satellite and is tested to ensure that all moving parts operate correctly in a vaccum, over a wide temp range. It is also subjected to vibration and shock testing to simulate vibration levels and or forces likely to be encountered on launch.

The thermal model contains all the electronics packages and other components that must be maintained at the correct temperature. Often the thermal, vaccum, and vibration tests of the entire satellite will be combined in a thermal vaccum chamber for what is known in the industry as a shake and bake test. The antennas are usually included on the thermal model to check for distortion of reflectors and displacement or bending of support structures.

The electrical model contains all the electronic parts of the satellite and is tested for correct electrical performance under total vaccum and a wide range of temperatures. the antennas of the electrical model must provide the correct beamwidth, gain & polarization properties.

Testing carried out on the prototype models is designed to overstress the system and induce failure in any weak components : temp cycling will be carried out to 10% beyond extremes, structural loads and G forces 50% above those expected in flight may be applied. Electrical equipment will be subjected to excess voltage and current drain to test for good electronic and thermal reliability.

Space qualification is an expensive process and one of the factors that makes large GEO satellites expensive. Some low earth orbit satellites have been built successively using less expensive techniques and relying on lower performance in orbit. Many of the electronic and mechanical components that are used in satellites are known to have limited life times, or a finite probability of failure. If failure of one of the components will reduce the comm. capacity of the satellite, a backup, or redundant, unit will be provided. The design of the system must be suchthat when one unit fails, the backup can automatically take over or be switched into operation by command from the ground.

Reliability :-

We need to be able to calculate the reliability of a satellite subsystem for two reasons

* we want to know what the probability is that the subsystem will still be working after a given time period.
* We need to provide redundant components or subsystems where the probability of a failure is too great to be accepted.

The manufactures of satellites must provide their customers with predictions of the reliability of the satellite and subsystems : to do this requires the use of reliability theory. Reliability theory is a mathematical attempt to predict the future and is therefore less certain than other mathematical techniques that operate in absolute terms. the application of reliability theory has enabled satellite engineers to build satellites that perform as expected at acceptable construction costs:-

The reliability of a component can be expressed in terms of the probability of failure after time t, $P_F(t)$. For most electronic equipment probability of failure is higher at the begening of life - the burn in period - than at some later time. As the component ages, failure becomes more likely, leading to the bathtub curve as shown in fig.

Semiconductors and IC that are required to have high reliability are subjected to burn in periods from 100 to 1000hours, often at a high temperature and excess voltage to induce failures in any suspect devices

The reliability of a device or subsystem is defined as

$$R(t) = \frac{N_s(t)}{N_0} = \frac{\text{no. of surviving components at time } t}{\text{no. of components at start of test period}}$$

The no. of components that failed in time t is $N_f(t)$

where $N_f(t) = N_0 - N_s(t)$.

Probability of any one of the N_0 components failing is related to the mean time before failure (MTBF). Suppose we continue testing devices until all of them fail. The ith device fails after time t_i where

$$MTBF = m = \frac{1}{N_0}\sum_{i=1}^{N_0} t_i$$

The average failure rate λ, is the reciprocal of the MTBF. If we assume that λ is a constant, then

$$\lambda = \frac{\text{no of failures in a given time}}{\text{no of surviving components}}$$

$$= \frac{1}{N_s}\frac{\Delta N_f}{\Delta t} = \frac{1}{N_s}\frac{dN_f}{dt} = \frac{1}{MTBF}$$

Failure rate λ is often given as the average failure rate per 10^9 h. The rate of failure, $\frac{dN_f}{dt}$ is the negative of the rate of survival $\frac{dN_s}{dt}$, so we can redefine λ as $\lambda = -\frac{1}{N_s}\frac{dN_s}{dt}$

$$\lambda R = \frac{N_s}{N_0} = \frac{-1}{N_0 R}\frac{d}{dt}(N_0 R) = \frac{-1}{R}\frac{dR}{dt}$$

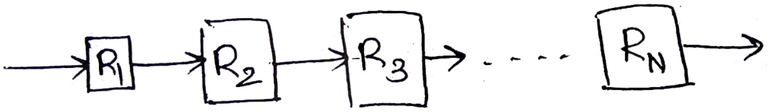
The sol of above eq is $R = e^{-\lambda t}$.

Thus the reliability of a device decreases exponentially with time, with zero reliability after infinite time i.e certain failure. End of useful life is usually taken to be the time t_1, at which R falls to $\frac{1}{e}(0.37)$ which is ~~the~~ when $t_1 = \frac{1}{\lambda} = m$. The probability of a device failing ∴ has an exponential relationship to the MTBF.

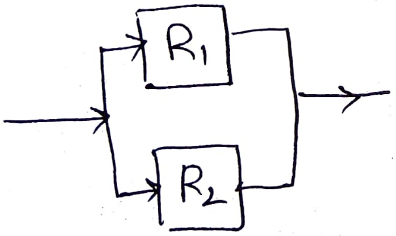
③

Redundancy :-

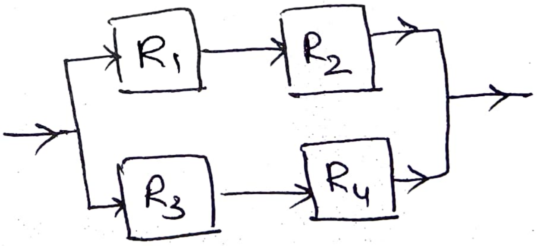
In a satellite, many devices are used, each with a different MTBF, and failure of one device may cause catastrophic failure of a complete subsystem. If we incorporate redundant devices, the subsystem can continue to function correctly. We can define three different situations for which we want to compute subsystem reliability: series connection, used in solar cells arrays, parallel connection, used to provide redundancy of the high power amplifiers in satellite transponders, and a switched connection, a series|parallel connection, widely used in electronic equipment.



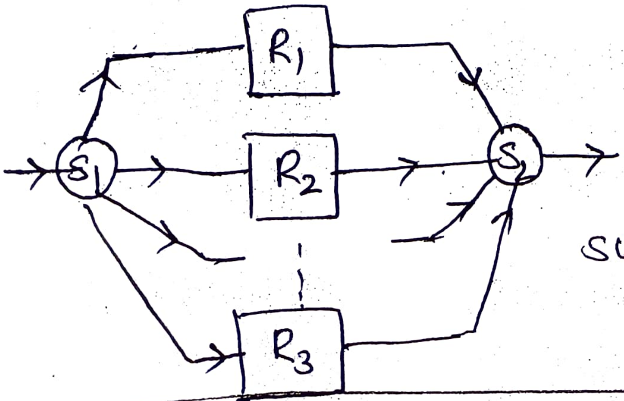
series connection



parallel connection



series|parallel connection



switched connection

The switched connection is also referred to as ring redundancy ~~of the high power amplifiers~~ since any component can be switched in for any other. The active devices (R_1, R_2, --- R_n) have sufficient bandwidth, power output range etc to be able to handle any of the channels that might be switched through to them. Most TWTAs and SSPAs are such wideband, large power range devices.

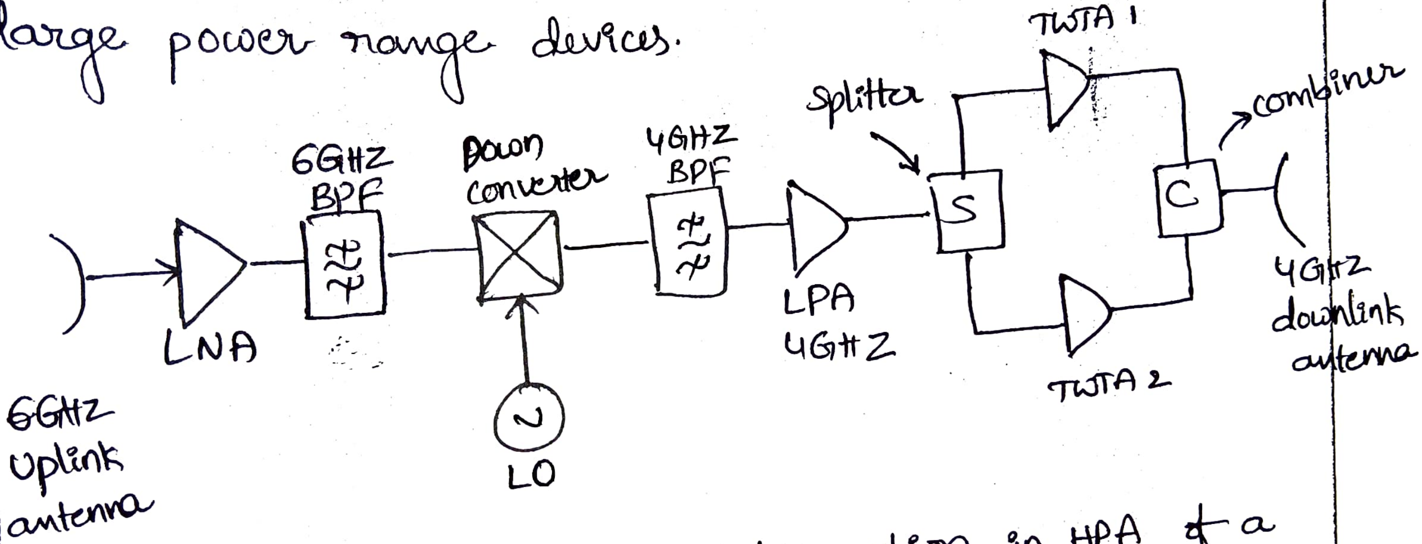


fig: Redundant TWTA configuration in HPA of a 6/4 GHZ bent pipe transponder

An example of parallel redundancy for the HPA is shown in fig. of a 6/4GHZ bent pipe transponder. The transponder has the high power output stage of the transponder has two parallel TWT amplifiers. If one TWTA fails, the other is switched on either automatically or by command from earth.

Unit - 3

Satellite Link Design

The design of a satellite comm. system is a complex process requiring compromises between many factors to achieve the best performance at an acceptable cost. The weight of a satellite is driven by two factors ; The number and output power of the transponders on the satellite and the weight of station keeping fuel three other factors infuence system design :

The choice of freq band
atmospheric propagation effects
multiple access technique.

The major bands are the 6/4 GHZ, 14/11 GHZ & 30/20GHZ bands. However, over much of the geostationary orbit there is already a satellite using both 6/4 GHZ and 14/11 GHZ every 2°. This is the minimum spacing used for satellites in GEO to avoid interference from uplink earth stations. Additional satellites can only be accomodated if they use another freq band such as 30/20 GHZ. Rain in the atmosphere attenuates radio signals. the effect is more severe as the freq increases. Attenuation through rain increases roughly as the square of freq. So a satellite uplink operating at 30GHZ suffers four times as much attenuation as an uplink at 14GHZ.

All comm. links are designed to meet certain performance objectives, usually a BER in a digital link or a signal to ratio S/N in an analog link measured in the baseband channel. The baseband channel BER or S/N ratio is determined by the C/N at the input to the demodulator in the receiver. Designing a satellite system therefore requires knowledge of the required performance of the uplink and downlink, the propagation characteristics and rain attenuation for the freq band, and the parameters of the satellite and the earth stations.

① Basic transmission theory :

The calculation of the power received by an earth station from a satellite transmitter is fundamental to the understanding of satellite communications. Two approaches are there for this : Use of flux density & link equation

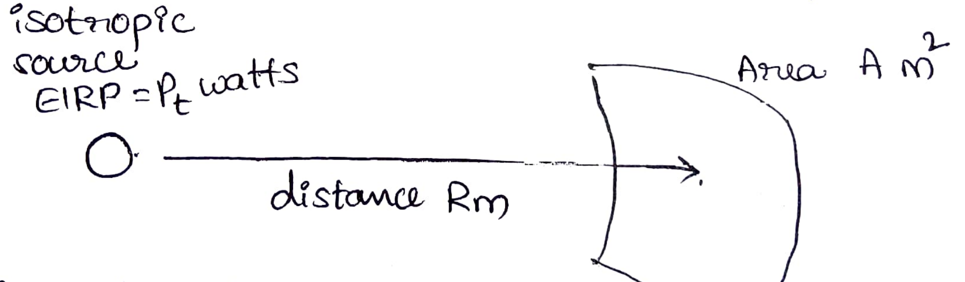


fig: flux density produced by an isotropic source

Consider a transmitting source, in free space, radiating a total power P_t watts uniformly in all directions. such a source is called isotropic. It is an idealization that cannot be realized physically because it could not create transverse EM waves. At a distance

R meters from the isotropic source transmitting RF power P_t watts, the flux density crossing the surface of a sphere with radius R is given by

$$F = \frac{P_t}{4\pi R^2} \quad W/m^2$$

All real antennas are directional and radiate more power in some directions than in others. Any real antenna has a gain $G(\theta)$, defined as the ratio of power per unit solid angle radiated in a direction θ to the average power radiated per unit solid angle

$$G(\theta) = \frac{P(\theta)}{P_0/4\pi}$$

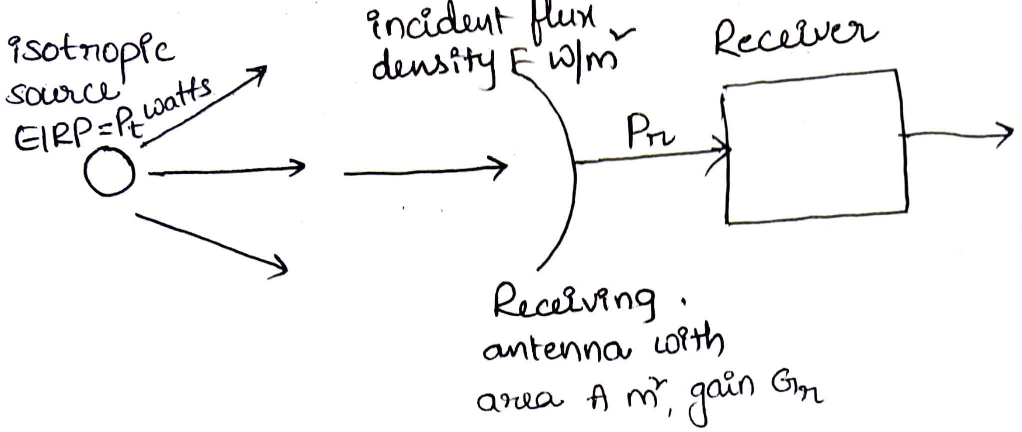
where $P(\theta)$ = power radiated per unit solid angle

P_0 = total power radiated

The reference for the angle θ is usually taken to be the direction in which maximum power is radiated, often called the boresight direction of the antenna. For a transmitter with output P_t watts driving a lossless antenna with gain G_t, the flux density in the direction of the antenna boresight at distance R meters is

$$F = \frac{P_t G_t}{4\pi R^2} \quad W/m^2$$

$P_t G_t$ = EIRP = it describes the combination of P_t & G_t in terms of an equivalent isotropic source with power $P_t G_t$, radiating uniformly in all directions.



If we had an ideal receiving antenna with an aperture area of A m^2 as shown in fig above, we would collect power P_r watts given by

$$P_r = F \times A \quad \text{watts}$$

A practical antenna with a physical aperture area of A_r m^2 will not deliver the power given in above eqn. Some of the energy incident on the aperture is reflected away from the antenna, and some is absorbed by lossy components. This reduction in efficiency is described by using an effective aperture A_e where $A_e = \eta_A A_r$ $\quad \eta_A$ = aperture efficiency

$\eta_A = 90\%$ for horn antenna

Thus the power received by a real antenna with a physical receiving area A_r and effective aperture area A_e m^2 is

$$P_r = \frac{P_t G_t A_e}{4\pi R^2} \text{ watts}$$

This eqn is essentially independent of freq if G_t and A_e are constant within a given band, the power received ~~by~~ at an earth station depends only on the EIRP of the satellite, the effective area of the earth station antenna, and the distance R.

The gain and area of an antenna are related by $G = \frac{4\pi A_e}{\lambda^2}$ where λ = wavelength

Sub ~~Ar~~ A_e in P_r eqn $\quad G = \frac{4\pi A_e}{\lambda^2} \Rightarrow A_e = \frac{G\lambda^2}{4\pi}$.

$$P_r = \frac{P_t G_t A_e}{4\pi R^2} = \cancel{\frac{P_t A_e}{4\pi R^2}\,\frac{4\pi A_e}{\lambda^2}}$$

$$= \frac{P_t G_t}{4\pi R^2}\left(\frac{G\lambda^2}{4\pi}\right) = \frac{P_t G_t G_r}{(4\pi R/\lambda)^2} \text{ Watts}$$

$$P_r = \frac{P_t G_t G_r}{\left(\frac{4\pi R}{\lambda}\right)^2}$$ This is known as link equation ①

and it is essential in the calculation of power received in any radio link. The term $\left(\frac{4\pi R}{\lambda}\right)^2$ is known as the path loss Lp.

$$\text{Power received} = \frac{\text{EIRP} \times \text{Receiving antenna gain}}{\text{path loss}} \text{ watts}$$

$$P_r = \text{EIRP} + G_r - L_p \text{ dbW}$$

where $\text{EIRP} = 10\log_{10}(P_t G_t)$ dbW

$$G_r = 10\log_{10}\left(\frac{4\pi A_e}{\lambda^2}\right) \text{ db}$$

$$L_p = 10\log_{10}\left(\frac{4\pi R}{\lambda}\right)^2 = 20\log\left(\frac{4\pi R}{\lambda}\right) \text{ db}$$

The above eqn $P_r = EIRP + G_r - L_p$ represents an idealized case, in which there are no additional losses in the link. In practice, we have losses in the atmosphere due to attenuation by oxygen, water vapor and rain, losses in the antennas at each end of the link. All of these factors are taken into account by the system margin but need to be calculated to ensure that the margin allowed is adequate.

$$P_r = EIRP + G_r - L_p - L_a - L_{ta} - L_{ra} \text{ dbw}$$

L_a = attenuation in atmosphere
L_{ta} = losses associated with Txing antenna
L_{ra} = " " " Rxing "

The received power P_r calculated in above eqns is commonly referred to as carrier power, C. This is because most satellite links use either freq modulation for analog Txion or phase modulation for digital Txion. In these modulation systems, the amplitude of the carrier is not changed when the data are modulated onto the carrier, so Rxed carrier power C is always equal to received power P_r

② ③

System noise temperature and $\frac{G}{T}$ ratio :

Noise temperature

It provides a way of determining howmuch thermal noise is generated by active and passive devices in the receiving system. At microwave frequencies a black body with a physical temperature T_p (kelvin) generates electrical noise over a wider bandwidth, The noise power is given by $P_n = KT_pB_n$

where K = Boltzmann's constant = 1.39×10^{-23} J/K

T_p = Physical temp of source (kelvin)

B_n = noise bandwidth in which the noise power is measured (Hz)

The term KT_p is a noise power spectral density in watts per hertz. The noise produced by the components of a low noise receiver. This can conveniently be done by equating the component to a black body radiator with an equivalent noise temperature, T_n Kelvin

In satellite comm. systems we are always working with weak signals and must make the noise level as low as possible to meet the C/N ratio requirements. This is done by making the bandwidth in the receiver, usually set by the IF amplifier stages, to be just large enough to allow the signal to pass unrestricted, while keeping the noise power to the lowest value possible.

To determine the performance of a receiving system we need to be able to find the total thermal noise power against which the signal must be demodulated. We do this by determining the system noise temperature T_s

If the overall end to end gain of the receiver is G_{rx} and its narrowest bandwidth is B_n HZ, the noise power at the demodulator input is

$$P_{no} = KT_s B_n G_{rx} \text{ watts.}$$

where G_{rx} = gain of the receiver from RF input to demodulator input.

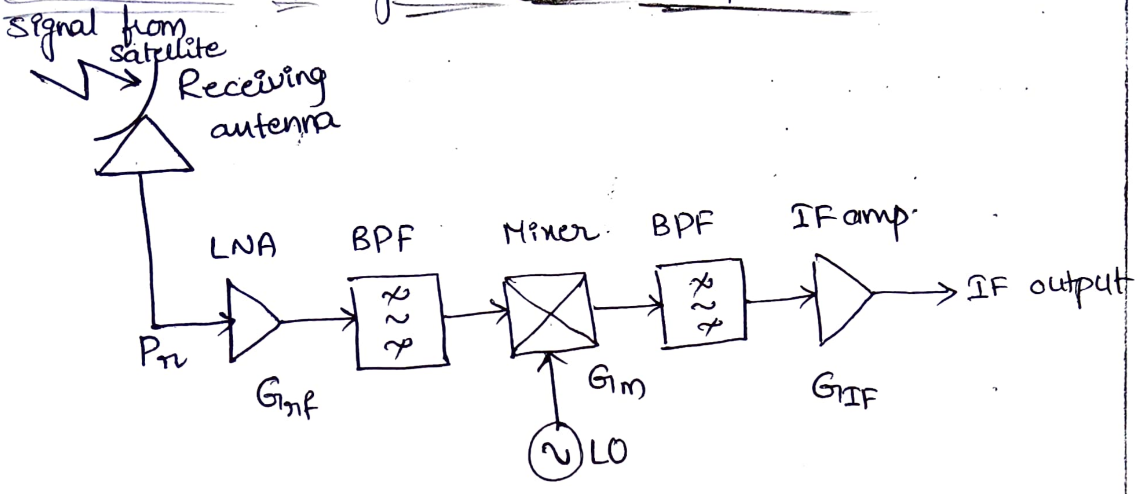
$P_n = KT_s B_n$ watts

= noise power referred to the input of the receiver

Let the antenna deliver a signal power P_r watts to the receiver RF input. The signal power at the demodulator input is $P_r G_{rx}$ watts, representing the power contained in the carrier and sidebands after amplification and freq conversion within the receiver. Hence, the carrier to noise ratio at the demodulator is given by

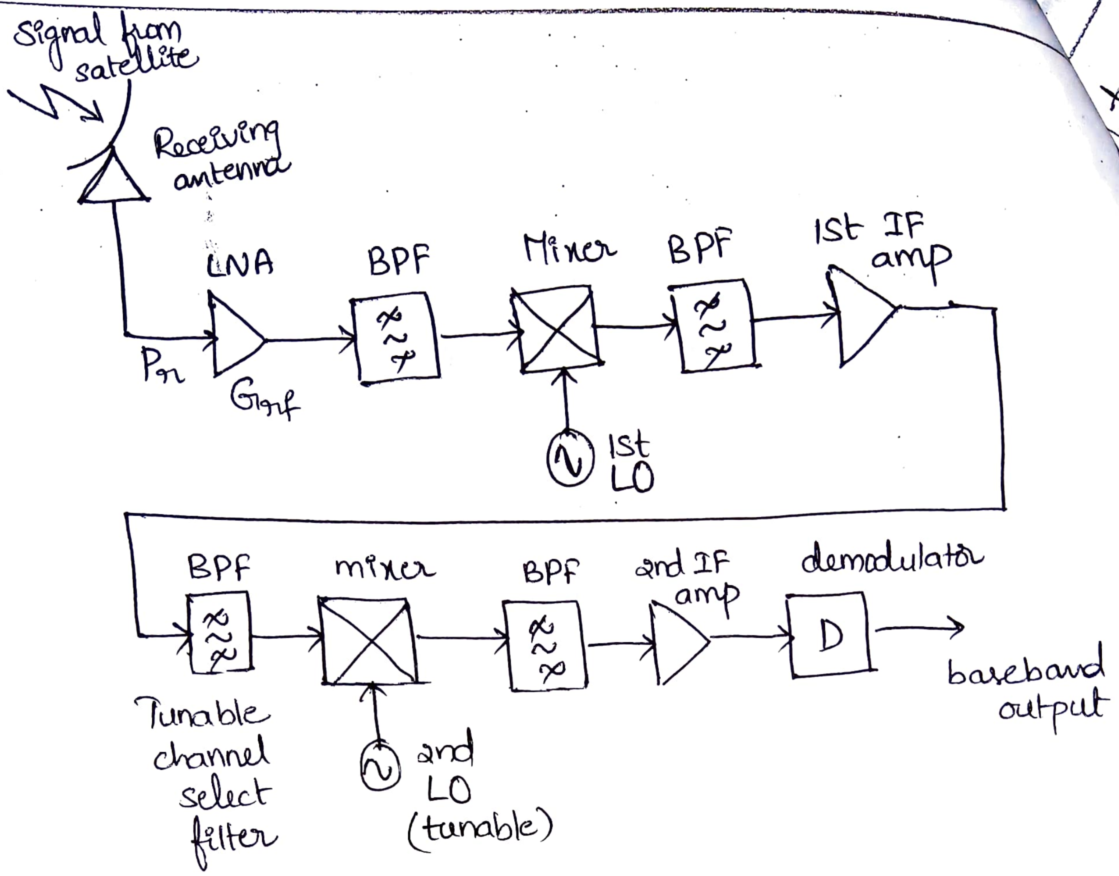
$$\frac{C}{N} = \frac{P_r G_{rx}}{KT_s B_n G_{rx}} = \frac{P_r}{KT_s B_n}$$

Calculation of system noise temperature :-



Above figure shows a simplified comm. receiver with an RF amplifier and single freq conversion, from its RF input to the IF output. This is the form used for all radio receivers, known as the superheterodyne. It has three main subsystems : a front end (RF amplifier, mixer and local oscillator) an IF amplifier (IF amp & filters) and a demodulator and baseband section.

The RF amplifier in a satellite communication receiver must generate as little noise as possible, so it is called a low noise amplifier or LNA. The mixer and LO form a freq conversion stage that downconverts the RF signal to a fixed IF, where the signal can be amplified and filtered accurately.



Many earth station receivers use the double superhet configuration shown in figure above, which has two stages of freq conversion. The front end of the receiver is mounted behind the antenna feed and converts the incoming RF signals to a first IF in the range 900 - 1400MHZ. The RF amplifier has a high gain and the mixer is followed by a stage of IF amplification. This section of the receiver is called a low noise block converter (LNB). The 900-1400MHZ signal is sent over a coaxial cable to set-top receiver that contains another down converter and a tunable local oscilator. The LO is tuned to convert the incoming signal from a selected transponder to a second IF freq. The 2nd IF amplifier has a bandwidth matched to the spectrum of the transponder signal.

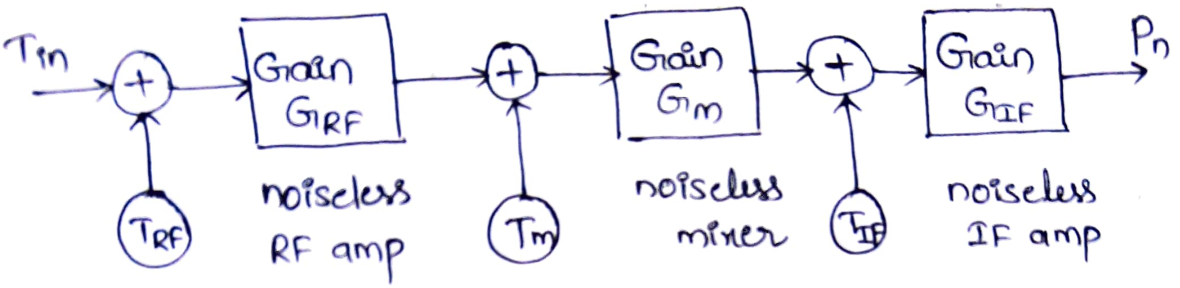
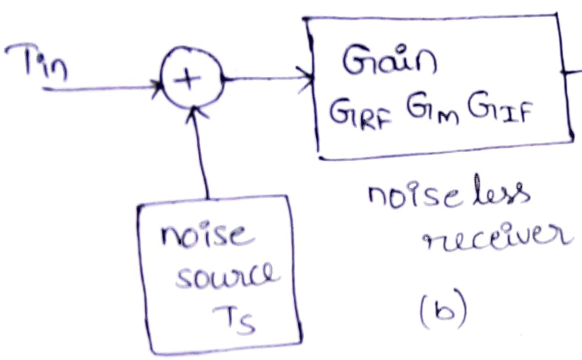
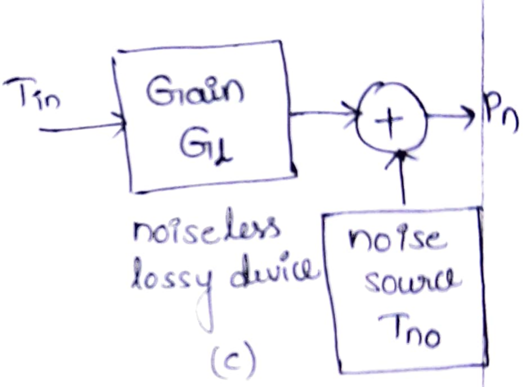


fig: Equivalent noise sources (a)



(b)



(c)

The equivalent circuits shown in fig can be used to represent a receiver for the purpose of noise analysis. The noisy devices in the receiver are replaced by equivalent noiseless blocks with the same gain and noise generators at the input to each block such that the block produces the same noise at its output as the device it replaces. The total noise power at the output of the IF amplifier of the receiver in fig is given by

$$P_n = G_{IF} K \cdot T_{IF} B_n + G_{IF} G_m K T_m B_m + G_{IF} G_m G_{RF} K B_n$$

where G_{RF} = Gain of RF amplifier

G_m = " " mixer

G_{IF} = " " IF amplifier

T_{RF}, T_m, T_{IF} = equivalent noise temperatures

T_{in} = noise temp of the antenna

Above eqn can be rewritten as

$$P_n = G_{IF} G_m G_{RF} \left[\frac{K T_{IF} B_n}{G_{RF} G_m} + \frac{K T_m B_n}{G_{RF}} + T_{RF} + T_{in} \right]$$

$$= G_{IF} G_m G_{RF} K B_n \left(T_{RF} + T_{in} + \frac{T_m}{G_{RF}} + \frac{T_{IF}}{G_{RF} G_m} \right)$$

The single source of noise shown in above fig with noise temperature T_s generates the same noise power P_n at its output if $P_n = G_{IF} G_m G_{RF} K T_s B_n$

The noise power at the output of the noise model in fig (b) will be the same as the noise power at the output of the noise model in fig(a) if

$$K T_s B_n = K B_n \left(T_{in} + T_{RF} + \frac{T_m}{G_{IF}} + \frac{T_{IF}}{G_m G_{RF}} \right)$$

Hence the equivalent noise source in fig (b) has a system noise temperature T_S where

$$T_S = T_{in} + T_{RF} + \frac{T_m}{G_{RF}} + \frac{T_{IF}}{G_m G_{RF}}$$

System noise temperature $T_S = T_{antenna} + T_{LNA}$

The noise model for an equivalent output noise source is shown in fig(c) and produces a noise temperature T_{no} given by $T_{no} = T_p(1 - G_l)$

where G_l = linear gain of the attenuating device or medium.

T_p = Physical temperature in kelvin of the device or medium.

Noise figure and noise temperature

Noise figure is used to specify the noise generated within a device. The noise figure is defined as

$$NF = \frac{(S/N)_{in}}{(S/N)_{out}}$$

Because noise temperature is more useful in satellite comm. systems, it is best to convert noise figure to noise temperature T_d. The relationship is $T_d = T_0(NF - 1)$

where T_0 = reference temp = 290K

$\frac{G}{T}$ ratio for earth stations :-

The link equation can be rewritten in terms of $\frac{C}{N}$ at the earth station

$$\frac{C}{N} = \left(\frac{P_t G_t G_r}{K T_s B_n}\right)\left(\frac{\lambda}{4\pi R}\right)^2 = \left(\frac{P_t G_r}{K B_n}\right)\left(\frac{\lambda}{4\pi R}\right)^2 \left(\frac{G_r}{T_s}\right)$$ ②

Thus $\frac{C}{N} \propto \frac{G_r}{T_s}$ and the terms in the square brackets are all constants for a given satellite system. The ratio $\frac{G_r}{T_s}$, which is usually quoted as simply $\frac{G}{T}$ in db with units db/K, can be used to specify the quality of a receiving earth station on a satellite receiving system, ∵ ↑ $\frac{G_r}{T_s}$ ⇒ ↑ $\frac{C}{N}$

Satellite terminals may be quoted as having a negative $\frac{G}{T}$ which is below 0 db/K. This means that the numerical value of G_r is lessthan the numerical value of T_s. ③

Design of downlinks :

The design of any satellit communication is based on two objectives

meeting a minimum $\frac{C}{N}$ ratio for a specified % of time

carrying the max revenue earning traffic at min cost

The art of good system design is to reach the best compromis of system parameters that meets the specification at less cost.

All satellite radio links are affected by rain attenuation. In the 6/4GHz band the effect of rain on the link is small. In the 14/11GHz (Ku) band, and even more so in the 30/20GHz (Ka) band, rain attenuation becomes all important. The C/N ratio in the receiver will fall below the min. permissible value for correct operation of the link for between 0.5 and 0.01% of the specified time, the link is then said to suffer an outage. Outages occur in heavy rain, usually in thunderstorms, and thunderstorm occurrences, rain [illegible] outage times of 0.1 to 0.5% of a year (9 to 44h) are usually tolerated in Ka band links. The allowable outage time for a link depends in part on the traffic carried.

Link budgets :-

$\frac{C}{N}$ ratio calculation is simplified by the use of link budgets. A link budget is a tabular method for evaluating the received power and noise power in a radio link. Link budgets use dB units for all parameters so that signal and noise powers can be calculated by addition and subtraction. Since it is usually impossible to design a satellite link at the first attempt, link budgets make the task much easier, because, once a link budget has been established, it is easy to change any of the parameters & recalculate the result.

The link budget must be calculated for an indiv transponder, and must be repeated for each of the individ links. In a two way satellite communication link there will be four separate links, each requiring a calculation of $\frac{C}{N}$ ratio. when a bent pipe transponder is used the uplink and downlink $\frac{C}{N}$ ratios must be combined to give an overall $\frac{C}{N}$.

Link budgets are usually calculated for a worst case, the one in which the link will have the lowest $\frac{C}{N}$ ratio. The calculation of $\frac{C}{N}$ ratio in a satellite link is based on the two equations for received signal power and receiver noise power. Received carrier power in db watts are

$$P_r = EIRP + G_r - L_p - L_a - \cancel{t} L_r - L_t \text{ dbW}.$$

A receiving terminal with a system noise temp T_s K and a noise bandwidth B_n Hz has a noise power P_n referred to the output terminals of the antenna where $P_n = K T_s B_n$ watts.

The receiving system noise power is usually written in db as $N = K + T_s + B_n$ dbW.

input is N_{xp} W where

$$N_{xp} = K + T_{xp} + B_n \text{ dbW}$$

where T_{xp} = system noise temperature of the transponder in dbK

B_n = dbHz

The power received at the input to the transponder is $P_{rxp} = P_t + G_t + G_r - L_p - L_{up}$ dbW

where $P_t G_t$ = uplink earth station EIRP in dbW

G_r = Satellite antenna gain in db in the direction of the uplink earth station

L_p = path loss in db.

L_{up} = All uplink losses otherthan pathloss

The value of $\left(\frac{C}{N}\right)_{up}$ at the LNA input of the satellite receiver is given by

$$\frac{C}{N} = 10 \log_{10}\left(\frac{P_r}{KT_sB_n}\right) = P_{rxp} - N_{xp} \text{ db}$$

The received power at the transponder input is also given by $P_{rxp} = N + \frac{C}{N}$ dbW.

The earth station transmitter output power P_t can also be calculated from the output power of the

Uplink design :

The uplink design is easier than the downlink in many cases, since an accurately specified carrier power must be presented at the satellite transponder and it is often feasible to use much higher power transmitters at earth stations than can be used on a satellite. The cost of transmitters tend to be high compared with the cost of receiving equipment in satellite comm. systems.

Earth station transmitter power is set by the power level required at the input to the transponder. This can be done in one of two ways. Either a specific flux density is required at the satellite, or a specific power level is required at the input to the transponder. Although flux density at the satellite is a convenient way to determine earth station transmit EIRP requirements, analysis of the uplink requires calculation of the power level at the input to the transponder so that the uplink $\frac{C}{N}$ ratio can be found. When a $\frac{C}{N}$ ratio is specified for the transponder, the calculation of required transmit power is straightforward. Let $(\frac{C}{N})_{up}$ be the specified $\frac{C}{N}$ ratio in the transponder, measured in a noise B.W B_n Hz. The noise power referred to the transponder

At frequencies above 10GHZ, (14GHZ & 30GHZ) propagation disturbances in the form of fading in rain cause the received power level at the satellite to fall. This lowers the uplink $\frac{C}{N}$ ratio in the transponder, which lowers the overall $\left(\frac{C}{N}\right)_0$ ratio in the earth station receiver when a linear (bent pipe) transponder is used. Uplink Power Control (UPC) can be used to combat uplink rain attenuation. Automatic monitoring and control of transmitted uplink power is used in 14GHZ uplink earth stations to maintain the uplink $\frac{C}{N}$ ratio in the satellite transponder during periods of rain attenuation.

Since the downlink is always at a different frequency from the uplink, a downlink attenuation of A_{db} must be scaled to estimate uplink attenuation. The scaling factor used is typically $\left(\frac{f_{up}}{f_{down}}\right)^a$ where a is typically between 2 & 2.4. The uplink attenuation is given by

$$A_{up} = A_{down}\left(\frac{f_{up}}{f_{down}}\right)^a$$

where A_{up} = Uplink rain attenuation

A_{down} = downlink " "

a = ~~2.2~~ 2 to 2.4

This uplink attenuation value applies only to rain and does not include gaseous attenuation.

transponder and transponder gain when these parameters are known and a bent pipe transponder is used. In general $P_{rxp} = P_{sat} - BO_0 - G_{xp}$ dbW

where P_{sat} = saturated power output of the transponder in dbW

BO_0 = output backoff in db

G_{xp} = gain of the transponder in db.

With small diameter earth stations, a higher power earth station transmitter is required to achieve a similar satellite EIRP. This has the disadvantage that the interference level at adjacent satellites rises, because, the, small earth station antennas has a wider beam. Thus it is not always possible to tradeoff transmitter power, aagainst uplink antenna size. There is a specification for transmit station antenna patterns, designed to minimize interference from adjacent uplinks. It is the uplink interference problem that determines satellite spacing and limits the capacity of the geostationary orbit in any freq band. To increase the capacity of the crowded geostationary orbit, intersatellite spacing could be reduced to 2°.

Uplink Power ~~&~~ Control cannot be applied until a certain amount of attenuation has built up in the link. As rain begins to affect the link between the earth station and satellite, the uplink $\frac{C}{N}$ ratio in the transponder will fall until upc starts to operate in the earth station transmitter.

Design for specified $\frac{C}{N}$:

The BER or $\frac{S}{N}$ ratio in the baseband channel of an earth station receiver is determined by the ratio of the carrier power to the noise power in the IF amplifier at the input to the demodulator. The noise present in the IF amplifier comes from many source. Till now we have considered only the receiver thermal noise and noise radiated by atmospheric gases and rain the slant path. when a complete satellite link is engineered, the noise in the earth station IF amplifier will have contributions from the receiver itself, the receiving antenna, sky noise, the satellite transponder from which it receives the signal, and adjacent satellites and terrestrial transmitters which share the same freq band.

When more than one $\frac{C}{N}$ ratio is present in the link, we can add the individual $\frac{C}{N}$ ratios reciprocally to obtain an overall $\frac{C}{N}$ ratio $\left(\frac{C}{N}\right)_o$.

The overall $\left(\frac{C}{N}\right)_o$ ratio is measured in the earth station at the output of the IF amplifier.

$$\left(\frac{C}{N}\right)_o = \frac{1}{\frac{1}{(C/N)_1} + \frac{1}{(C/N)_2} + ----}$$

this is sometimes referred to as the reciprocal $\frac{C}{N}$ formula. The $\frac{C}{N}$ values must be linear ratios, not decibel values. Since the noise power in the individual $\frac{C}{N}$ ratios is referenced to the carrier power at that point, all the c values in above eqn are same.

$$\therefore \left(\frac{C}{N}\right)_o = \frac{1}{\frac{N_1}{C} + \frac{N_2}{C} + \cdots\cdots} = \frac{C}{N_1 + N_2 + N_3 + ---}$$

In decibel units

$$\left(\frac{C}{N}\right)_o = C\,dbW - 10\log_{10}(N_1 + N_2 + ---- W)\ db$$

Note that $\left(\frac{C}{N}\right)_{dn}$ cannot be measured at the receiving earth station. $\frac{C}{N}$ ratio measurement at the receiver will always yield $\left(\frac{C}{N}\right)_o$, the combination of transpon and earth station $\frac{C}{N}$ ratios.

To calculate the performance of a satellite link we must therefore determine the uplink $(\frac{C}{N})_{up}$ ratio in the transponder and the downlink $(\frac{C}{N})_{dn}$ in the earth station receiver. We must also consider whether there is any interference present, either in the satellite receiver or the earth station receiver. If the Intermodulation products (IM) power level in the transponder is known, a $\frac{C}{I}$ value can be found and included in the calculation of $(\frac{C}{N})_o$ ratio. Interference from adjacent satellites is likely whenever small receiving antennas are used, as with VSATs (Very small aperture terminals) and DBS-TV receivers.

Overall $(\frac{C}{N})_o$ with uplink and downlink attenuation;

Most satellite links are designed with link margins to allow for attenuation that may occur in the link or increases in noise power caused by interference. The effect of a change in the uplink $\frac{C}{N}$ ratio has a different impact on overall $(\frac{C}{N})_o$ depending on the operating mode and gain of the transponder. There are 3 different transponder types or operating modes.

Linear transponder : $P_{out} = P_{in} + G_{xp}$ dbW

Nonlinear " : $P_{out} = P_{in} + G_{xp} - \Delta G$ dbW

Regenerative " : $P_{out} =$ constant dbW

where P_{in} = power delivered by the satellites recei
antenna to the input of the transponder

P_{out} = Power delivered by the transponder HPA
to the input of satellites T_x antenna.

G_{xp} = Gain of the transponder.

The maximum output power from a transponder is called the saturated output power and is the nominal transponder power output rating that is usually quoted. The transponder input output characteristic is highly nonlinear when operated at this output power level. When a transponder is operated close to its saturated output power level, digital waveforms are changed, resulting in intersymbol interference, and FDMA operation results in the generation of intermodulation products by multiplication of the individual signals. Transponders are usually operated with output backoff to make the characteristics more nearly linear. The exact amount of output backoff required in any given application depends on the characteristics of the transponder and the signals it carries. Typical values of output backoff are 1db for a single FM or PSK carrier to 3db for FDMA operation with several carrier.

Uplink and downlink attenuation in rain ;

Rain attenuation affects uplink and downlinks differently. We usually assume that rain attenuation is occuring on either the uplink or the downlink, but not on both at the same time. This is usually true for earth stations that are well separated geographically, but not if they are close together (<20km). Heavy rain occurs with a somewhat random geographic distribution for lessthan 1% of time, so the probability of significant attenuation occuring on both the uplink and downlink simultaneously is small. In the following analysis of uplink and downlink attenuation effects, it will be assumed that one link is attenuated and the other is operating in clear air

Uplink attenuation and $\left(\frac{C}{N}\right)_{up}$

The transponder receiver noise temperature does not change significantly when rain in present in the uplink path to the satellite. The satellite receiving antenna beam is always sufficiently wide that it sees a large area of the earth's surface and local noise temperature variations are insignificant because the satellite antenna beam sees the tops of cumulonimbus clouds above the rain, instead of the earth's surface.

Rain attenuation on the uplink path to the satellite reduces the power at the satellite receiver input, and thus reduces $\left(\frac{C}{N}\right)_{up}$ in direct proportion to the attenuation on the slant path. If the transponder is operating in a linear mode, the output power will be reduced by the same amount, which will cause $\left(\frac{C}{N}\right)_{dn}$ to fall by an amount equal to the attenuation on the uplink. Hence for the case of a linear transponder and rain attenuation in the uplink of A_{up} db

$$\left(\frac{C}{N}\right)_{o\ \text{uplink rain}} = \left(\frac{C}{N}\right)_{o\ \text{clear air}} - A_{up}\ \text{db} \quad \text{linear transponder}$$

If the transponder is nonlinear, the reduction in input power caused by uplink attenuation of A_{up} db results in a smaller reduction in output power, by an amount ΔG

$$\left(\frac{C}{N}\right)_{o\ \text{uplink rain}} = \left(\frac{C}{N}\right)_{o\ \text{clear air}} - A_{up} + \Delta G\ \text{db} \quad \text{nonlinear transponder}$$

If the transponder is digital and regenerative or incorporates an automatic gain control (AGC) system to maintain a constant output power level

$$\left(\frac{C}{N}\right)_{o\ \text{uplink rain}} = \left(\frac{C}{N}\right)_{o\ \text{clear air}}\ \text{db} \quad \text{regenerative transponder}$$

The above eqn will hold only if the received signal is above threshold and the BER of the recovered signal in the transponder is small.

Downlink attenuation and $\left(\frac{C}{N}\right)_{dn}$

The earth station receiver noise temperature can change very significantly when rain is present in the downlink path from the satellite. The sky noise temperature can increase to close to the physical temperature of the individual rain drops, particularly in very heavy rain. The result is that the received power level, C, is reduced and the noise power, N, in the receiver increases. The result for downlink $\frac{C}{N}$ is given by

$$\left(\frac{C}{N}\right)_{dn\ rain} = \left(\frac{C}{N}\right)_{dn\ clear\ air} - A_{rain} - \Delta N_{rain} \quad db$$

The overall $\frac{C}{N}$ is then given by

$$\left(\frac{C}{N}\right)_0 = \frac{1}{\frac{1}{(C/N)_{dn\ rain}} + \frac{1}{(C/N)_{up}}} \quad db$$

System design for specific performance :-

A typical two way satellite communication link consists of four separate paths ; an out bound uplink path from one terminal to the satellite and an outbound downlink to the second terminal, and an inbound uplink from the 2nd terminal to the satellite an inbound downlink to the first terminal. The links in the two directions are independent and can be designed separately, unless they share a single transponder using FDMA. A broadcast link, like the DBS-TV system is a one way system, with just one uplink and one downlink.

Satellite communication link design procedure

The design procedure for a one way satellite communication link can be summarized by the following 10 steps. The return link design follows the same procedure.

1. Determine the freq band in which the system must operate. Comparative designs may be required to help make the selection.
2. Determine the communications parameters of the satellite. Estimate any values that are not known.
3. Determine the parameters of the transmitting and receiving earth stations.
4. Start at the transmitting earth station. Establish an uplink budget and a transponder noise power budget to find $(\frac{C}{N})_{up}$ in the transponder.
5. Find the output power of the transponder based on transponder gain or output backoff.
6. Establish a downlink power and noise budget for the receiving earth station. Calculate $(\frac{C}{N})_{dn}$ and

$\left(\frac{C}{N}\right)_0$ for a station at the edge of the coverage zone. (worst case)

7. Calculate $\frac{S}{N}$ or BER in the baseband channel. Find the link margins.

8. Evaluate the result and compare with the specification requirements. Change parameters of the system as required to obtain acceptable $\left(\frac{C}{N}\right)_0$ a $\frac{S}{N}$ or BER values. This may require several trial designs.

9. Determine the propagation conditions under which the link must operate. Calculate outage times for the uplinks and downlinks.

10. Redesign the system by changing some parameters if the link margins are inadequate. Check that all the parameters are reasonable, and that the design can be implemented within the expected budget.

Multiple Access

The ability of the satellite to carry many signals at the same time is known as multiple access. Multiple access allows the communication capacity of the satellite to be shared among a large no. of earth stations. The basic form of multiple access employed by all communications satellites is the use of many transponders. A large GEO satellite may have a communication band width of over 200MHZ with in an allocated spectrum of 500MHZ. Through frequency reuse with multiple antenna beams and orthogonal polarization, the spectrum can be reused several times. The freq spectrum used by the satellite is divided into smaller bandwidths which are allocated to transponders, allowing separate comm. links to be established via the satellite on the basis of transmit freq. Transponder bandwidths of 36, 54 & 72MHZ have been commonly employed on GEO satellites.

The signals that earth stations transmit to a satellite may differ widely in their character - voice, video, data, facsimile - but they can be sent through the same satellite using multiple access and multiplexing techniques. Multiplexing is the process of combining a no. of signals into a single signal, so that it can be processed by a single amplifier or transmitted over a single radio channel.

The designer of a satellite communication system must make decisions about the form of multiple access to be used. The multiple access technique will influence the capacity and flexibility of the satellite comm. system, its cost, and its ability to earn revenue. There are three basic multiple access techniques. In FDMA all users share the satellite at the same time, but each user transmits at a unique allocated freq. FDMA can be used with analog or digital signals. In TDMA each user is allocated a unique timeslot at the satellite so that signals pass through the transponder sequentially. Because TDMA causes delays in transmission, it is used only with digital signals. In code division multiple access (CDMA) all users transmit to the satellite on the same frequency and at the same time. The earth stations transmit orthogonally coded spread spectrum signals that can be separated at the receiving earth station by correlation with the transmitted code. CDMA is inherently a digital technique.

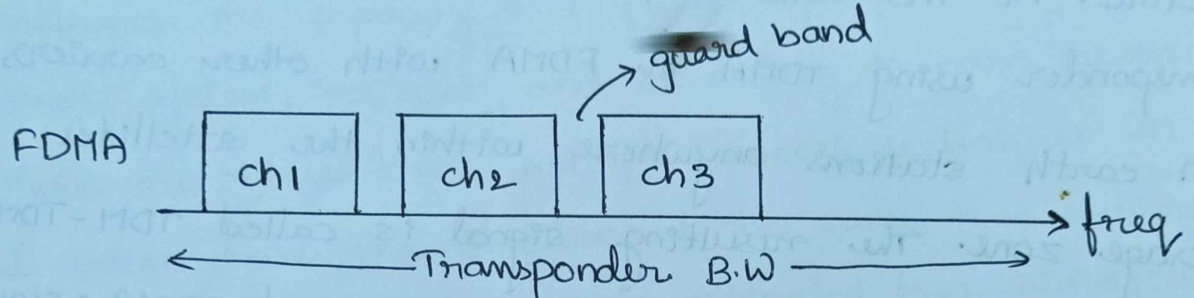
The distinction between multiplexing and multiple access is, multiplexing applies to signals that are generated at one location, where as multiple access refers to signals from a no. of different geographical locations.

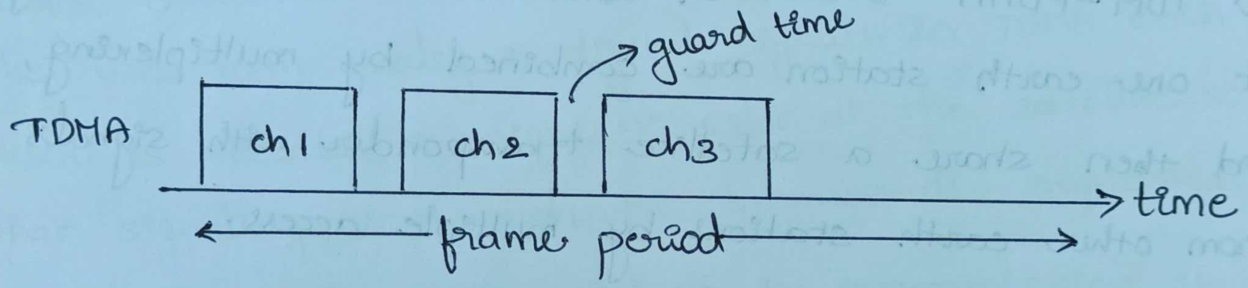
For example, an earth station might use time division multiplexing (TDM) to create a high speed digital data stream from many digital speech channels delivered to that earth station, and then modulate the data stream onto an RF carrier and transmit the carrier to the satellite. At the satellite, the carrier can share a transponder using TDMA or FDMA with other carriers from earth stations anywhere within the satellite's coverage zone. The resulting signal is called TDM-TDMA or TDM-FDMA. Distinction between TDM and TDMA: signals at one earth station are combined by multiplexing, and then share a satellite transponder with signals from other earth stations by multiple access.

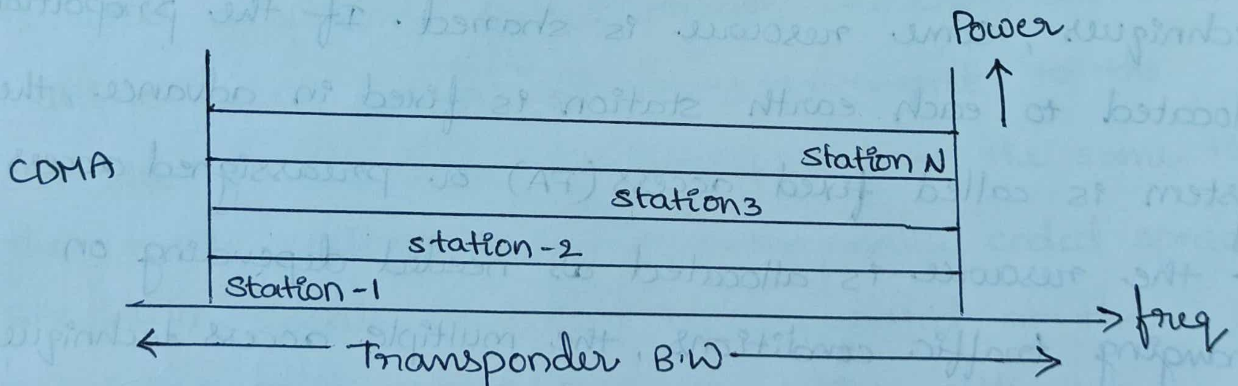
In all three of the classic multiple access techniques, some resource is shared. If the proportion allocated to each earth station is fixed in advance, the system is called fixed access (FA) or preassigned access (PA) If the resource is allocated as needed depending on changing traffic conditions, the multiple access technique is called demand access (DA). Demand access with FDMA is widely used in VSAT systems. Fixed assignment would be wasteful of transponder capacity, so demand assignment is used.

Systems which combine both FDMA and TDMA techniques are sometimes called hybrid multiple access schemes or multifrequency TDMA (MF-TDMA).

Frequency Division Multiple Access (FDMA) :







FDMA was first multiple access technique used in Satellite comm. systems. when SC began in the 1960's, most of the traffic carried by satellites was telephony. All signals were analog, and analog multiplexing was used at earth stations to combine large no of telephone channels into a single baseband signal that could be modulated onto a single RF carrier. Individual telephone channels can be shifted in freq from baseband to a higher freq so that they can be stacked into a group of channels using FDM.

Early satellite systems used FDM to multiplex upto 1800 telephone channels into a wide baseband occupying upto 8MHZ, which was modulated onto an RF carrier using frequency modulation (FM). The FDM-FM RF carrier was transmitted to the satellite, where it shared a transponder with other carriers using FDMA. The technique is known as FDM-FM-FDMA, and was the preferred method for the transmission of telephone channels over Intelsat satellites for more than 20 years. The main advantage of FDMA is that filters can be used to separate signals. Microwave filters were used in earth stations to separate the FDMA signals within a given transponder.

The used of MW filters to separate channels made the fixed assignment approach to FDMA very inflexible changing the freq assignment or bandwidth of any one transmitting earth station required returning of the MW filters at several receiving earth stations.

With fixed assignment, the frequencies and satellite capacity cannot be reallocated between routes, so much of the satellite capacity remains idle. Estimates of average loading of Intelsat satellites using fixed assignment are typically around 15%. Demand assignment and single channel per carrier (SCPC) techniques allow higher loadings and therefore give satellite operators increased revenue.

Every earth station that operates in an FDMA network must have a separate IF receiver for each of the carriers that it wishes to receive. SCPC systems can have a very large no. of carriers in one transponder, as a result, FDMA earth stations ~~tend~~ tend to have a very large no. of no of IF receivers and demultiplexers which select individual carriers using narrowband IF filters. Guard bands are essential in FDMA systems to allow the filters in the receivers to select individual channels without excessive interference from adjacent channels. Typically guard bands of 10 to 25% of the channel bandwidth are needed to minimize adjacent channel interference.

FDM-FM-FDMA was a telephone transmission technique well suited to analog telephone signals. Telephony has largely become digital and FDM has been replaced by TDM. Digital speech is now used throughout telephone systems, so multiple telephone channels are always transmitted as a high speed digital signal.

FDMA is widely used as a method of sharing the bandwidth of satellite transponders. Ideally, a satellite would carry a very large no.of transponders, each of which could be allocated to a single RF carrier. In the case of telephony, each transponder would have a B.W exactly matched to the RF spectrum of the transmitted telephone channel, with tight filtering to ensure that each signal can be separated from adjacent signals. This approach is impractical; thousands of transponders would be needed and the satellite could be used only for telephony. The builders & operators of satellites have historically shown a strong preference for wideband transponders that can carry any type of traffic - the bent pipe transponder that can carry voice, video or data. As a result, transponders have always had wide bandwidths, with B.W of 36, 54 & 72MHZ commonly employed. When an earth station has a carrier that occupies less than the transponder B.W, FDMA can be used to allow that carrier to share the transponder with other carriers.

Allocating a wideband transponder to a single narrow bandwidth signal is clearly wasteful, so FDMA is a widely used technique. When an earth station sends one signal on a carrier, the FDMA access technique is called single channel per carrier (SCPC). Thus a system in which a large no.of small earth stations, such as mobile telephones, access a single transponder using FDMA is called a SCPC-FDMA. Hybrid multiple access schemes can use TDM of baseband channels which are then modulated onto a single carrier. A no. of earth stations can share a transponder using FDMA, giving a system known as TDM-SCPC-FDMA. Seq of abbreviations is baseband multiplexing technique first, then multiple access technique next. TDM-SCPC-FDMA is used by VSAT networks.

FDMA has a disadvantage in SC systems when the satellite transponder has a nonlinear characteristics. Most satellite transponders use HPA which are driven close to saturation, causing nonlinear operation. A transponder using TWT amplifier (TWTA) is more prone to nonlinearity than one with a solid state high power amplifier (SSHPA). Equalization at the transmitting station, can sometimes be employed to linearize the transponder when fixed assignment is used. Linearization of solidstate and TWT HPAS on the satellite is also possible. Nonlinearity of the transponder HPA causes a reduction in the overall $\left(\frac{C}{N}\right)_0$ ratio at the receiving earth station when FDMA is used because intermodulation (IM) products are generated in the transponder. Some of the IM products will be within the transponder bandwidth and will cause interference.

Intermodulation

IM products are generated whenever morethan one signal is carried by a nonlinear device. Sometimes filtering can be used to remove the IM products, but if they are with in the B.W of the transponder they cannot be filtered out. The saturation characteristic of a transponder can be modeled by a cubic curve to illustrate the generation of 3rd order intermodulation. 3rd order IM is important because 3rd order IM products often have frequencies close to the signals that generate the intermodulation, and therefore likely to be within the transponder B.W.

To illustrate the generation of 3rd order IM products, we will model a nonlinear characteristic of the transponder HPA with a cubic voltage relationship and apply to unmodulated carriers at freq f_1 and f_2 at the input of the amplifier.

$$V_{out} = A V_{in} + b(V_{in})^3 \quad \text{where } A >> b.$$

The amplifier input signal is $V_1 \cos\omega_1 t + V_2 \cos\omega_2 t$

The amplifier output signal is $V_{out} = A V_{in} + b(V_{in})^3$

$$V_{out} = \underbrace{A(V_1 \cos\omega_1 t + V_2 \cos\omega_2 t)}_{\text{linear term}} + \underbrace{b(V_1 \cos\omega_1 t + V_2 \cos\omega_2 t)^3}_{\text{cubic term}}$$

The linear term simply amplifies the input signal by a voltage gain A. The cubic term, denoted as V_{3out} can be expanded as

$$V_{3out} = (V_1 \cos\omega_1 t + V_2 \cos\omega_2 t)^3 \cdot b$$

$$= b\left(V_1^3 \cos^3\omega_1 t + V_2^3 \cos^3\omega_2 t + 2V_1^2 \cos^2\omega_1 t \cdot V_2 \cos\omega_2 t + 2V_2^2 \cos^2\omega_2 t \cdot V_1 \cos\omega_1 t\right)$$

The 1st two terms contain frequencies f_1, f_2, $3f_1$ & $3f_2$. The triple freq components can be removed from the amplifier output with a B.P.F. The 2nd two terms generate the 3rd order IM freq components.

We can expand $\cos^2 x = \frac{1}{2}(\cos 2x + 1)$. Hence the IM terms of interest become

$$V_{IM} = 2bV_1^2 \cos^2\omega_1 t \cdot V_2 \cos\omega_2 t + 2bV_2^2 \cos^2\omega_2 t \cdot V_1 \cos\omega_1 t$$

$$= \cancel{2}bV_2 \cos\omega_2 t \cdot V_1^2 \frac{1}{\cancel{2}}(\cos 2\omega_1 t + 1) + \cancel{2}bV_1 \cos\omega_1 t \cdot V_2^2 \frac{1}{\cancel{2}}(\cos 2\omega_2 t + 1)$$

$$= bV_1^2 V_2 \cos\omega_2 t \cdot (\cos 2\omega_1 t + 1) + bV_2^2 V_1 \cos\omega_1 t\,(\cos 2\omega_2 t + 1)$$

$$= bV_1^2 V_2 (\cos\omega_2 t \cdot \cos 2\omega_1 t + \cos\omega_2 t) + bV_2^2 V_1 (\cos\omega_1 t \cdot \cos 2\omega_2 t + \cos\omega_1 t)$$

The terms at frequencies f_1 and f_2 add to the wanted output of the amplifier, so the 3rd order IM products are generated by the $f_1 \times 2f_2$ and $f_2 \times 2f_1$ terms.

Using $\cos x \cdot \cos y = \cos(x+y) + \cos(x-y)$

The output of the amplifier contains IM freq components given by

$$V'_{IM} = bV_1^2 \cdot V_2 \left[\cos(2\omega_1 t + \omega_2 t) + \cos(2\omega_1 t - \omega_2 t)\right]$$
$$+ bV_2^2 \cdot V_1 \left[\cos(2\omega_2 t + \omega_1 t) + \cos(2\omega_2 t - \omega_1 t)\right]$$

We can filter out the sum terms in above eqn, but the difference terms, with frequencies $2f_1 - f_2$, $2f_2 - f_1$ may fall within the transponder bandwidth. These two terms are known as the 3rd order intermodulation products and are given by V_{3IM} where

$$V_{3IM} = bV_1^2 V_2 \cos(2\omega_1 t - \omega_2 t) + bV_2^2 V_1 \cos(2\omega_2 t - \omega_1 t)$$

The magnitude of the IM products depends on the parameter b, which describes the nonlinearity of the transponder, and the magnitude of the signals. The wanted signals at the transponder output, at freq f_1 & f_2 have magnitudes AV_1 and AV_2. The wanted o/p at amplifier is

$$V_{out} = AV_1 \cos\omega_1 t + AV_2 \cos\omega_2 t$$

The total power of the wanted output from the HPA, referenced to a 1Ω load, is ∴

$$P_{out} = \frac{1}{2}A^2V_1^2 + \frac{1}{2}A^2V_2^2 = A^2(P_1 + P_2)W$$

where P_1 & P_2 are the power levels of the wanted signals. The power of the IM products at the output of the HPA is

$$P_{IM} = 2\left(\frac{1}{2}b^2V_1^6 + \frac{1}{2}b^2V_2^6\right)$$

$$= b^2\left(P_1^3 + P_2^3\right) W$$

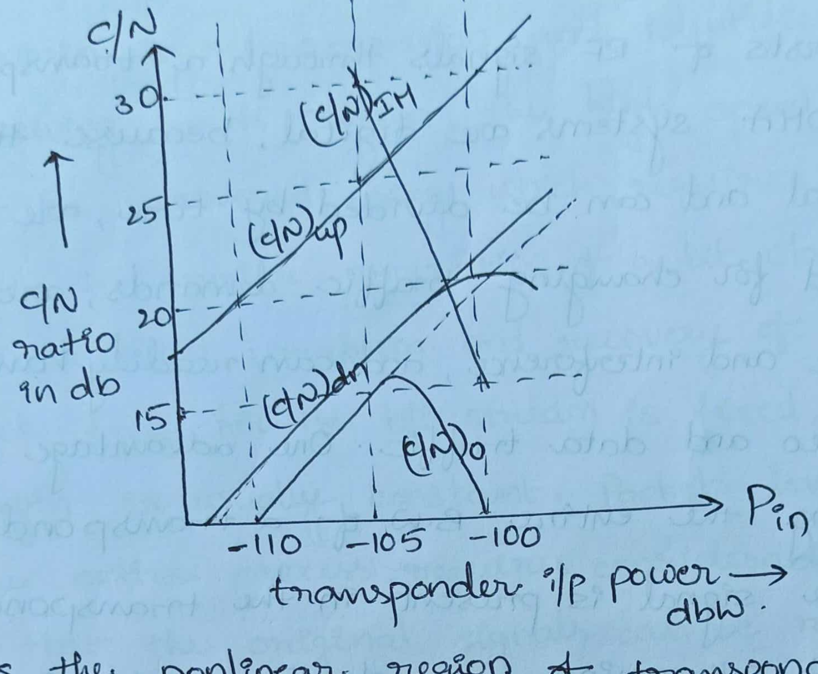
IM products ↑ in proportion to the cubes of the signal powers, with power levels that depend on the ratio $\left(\frac{b}{A}\right)^2$. The greater the nonlinearity, the larger the IM products $\left(\frac{b}{A}\uparrow\right)$.

Calculation of $\frac{C}{N}$ with intermodulation

Intermodulation between carriers in a nonlinear transponder adds unwanted products into the transponder B.W that are treated as though the interference were gaussian noise. The output backoff of a transponder reduces the $\frac{C}{N}$ ratio in the transponder

$$\left(\frac{C}{N}\right)_0 = \frac{1}{\frac{1}{(C/N)_{up}} + \frac{1}{(C/N)_{dn}} + \frac{1}{(C/N)_{IM}}}$$

There is an optimum output backoff for any linear transponder operating in FDMA mode. Figure illustrates the effect of the HPA operating point on each $\frac{C}{N}$ ratio in above eqn. when the operating point is set by the power transmitted by the uplink earth station. The uplink $\left(\frac{C}{N}\right)_{up}$ ratio ↑ linearly as the transponder i/p power ↑, leading to a corresponding nonlinear ↑ in transponder output power.



As the nonlinear region of transponder is reached, the downlink $\left(\frac{C}{N}\right)_{dn}$ ratio ↑ less rapidly than $\left(\frac{C}{N}\right)_{up}$ because the nonlinear transponder is going into saturation IM products start to appear as the nonlinear region is approached, increasing rapidly as saturation is reached. With a 3rd order model for nonlinearity, the intermodulation products ↑ in power at 3times the rate at which the i/p power to the transponder is increases, causing $\left(\frac{C}{N}\right)_{IM}$ to ↓ rapidly as saturation is approached. The overall $\left(\frac{C}{N}\right)_{0}$ ratio in the receiving earth station receiver has a max value at an i/p power level of -104dbW in the example in above fig. This is the optimum operating point for the transponder

Time division multiple access (TDMA)

In TDMA a no.of earth stations take turns transmitting bursts of RF signals through a transponder. All practical TDMA systems are digital, because the signals are digital and can be divided by time, are easily reconfigured for changing traffic demands, are resistant to noise and interference, and can readily handle mixed voice, video and data traffic. One advantage of TDMA when using the entire B.W of a transponder is that only one signal is present in the transponder at one time, thus overcoming many of the problems caused by nonlinear transponders operating with FDMA. However using all of the transponder B.W requires every earth station to transmit at a high bit rate, which requires high transmitted power, and TDMA is not well suited to narrow band signals from small earth stations. Nonlinearity in the transponder can cause an ↑ in ISI with digital carriers, equalizers can be used at the receiving earth stations to mitigate the effect.

The difference between TDM and TDMA is that TDM is ~~tha~~ a baseband technique used at one location (a $T_x$ing earth station) to multiplex several digital bit streams into a single higher speed digital signal.

Group of bits are taken from each of the bit streams and formed into baseband packets or frames that also contain synchronization and identification bits. At a receiving earth station, the high speed bit stream must 1st be recovered, ~~using~~ which requires demodulation of the RF carrier, generation of a bit clock, sampling of the received waveform and recovery of the bits. The clock freq for the bit stream is fixed, and the frame length is usually constant, Packet lengths can vary. The entire process requires considerable storage of bits so that the original signals can be rebuit, leading to delays in transmission. In a GEO satellite system, the largest delay is always the transmission time to the satellite and back to earth, typically 240ms. The transmission delay is unavoidable, but any additional delays should be minimized.

In a TDMA system, the RF carrier from each earth station sharing a transponder is sent as a burst at a specific time. At the satellite, bursts from different earth stations arrive sequentially, so the transponder carries a near continuous signal made up of a seq of short bursts coming from different earth stations.

The frame has a length from 125us to many (ms) milli seconds, and the burst from the earth station must be transmitted at the correct time to arrive at the satellite in the correct position within the TDMA frame. This requires synchronization of all the earth stations in a TDMA network, adding complexity to the transmitting station. Each station must know exactly when to transmit, typically with in a milli second, so that the RF bursts arriving at the satellite from different earth stations do not overlap (overlap → collision, not allowed in TDMA)

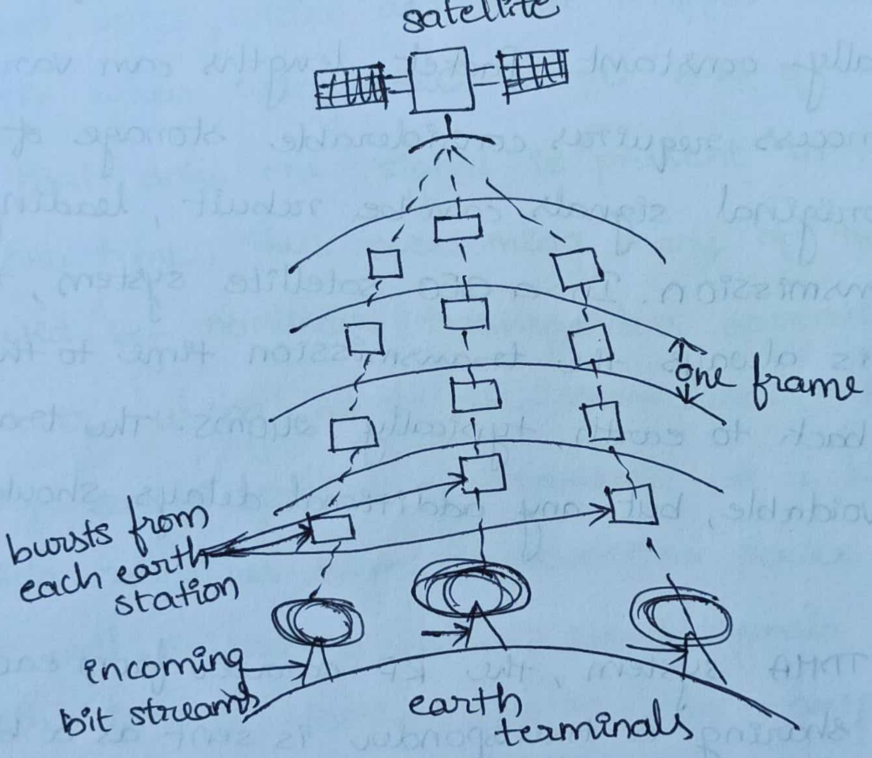
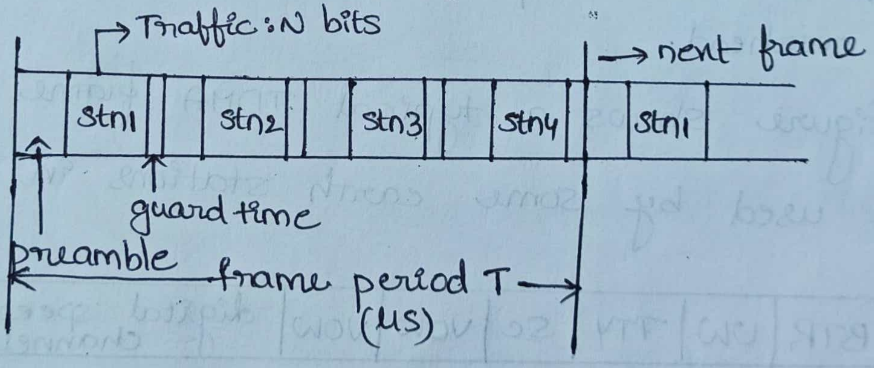


fig: illustration of TDMA with three earth stations

A receiving earth station must synchronize its receiver to each of the sequential bursts in the TDMA signal and recover the transmission from each uplink earth station. The uplink transmissions are then broken down to ~~exact~~ ~~th~~ extract the data bits, which are stored and reassembled into their original bit streams for onward transmission. The individual transmissions from different uplink earth stations are usually sent using BPSK or QPSK and have small differences in carrier and clock frequencies and different carrier phases.

TDMA frame structure

A TDMA frame contains the signals transmitted by all of the earth station in a TDMA network. It has a fixed length, and is built up from the burst transmissions of each earth station, with guard times between each burst. Figure shows a simplified diagram of a TDMA frame for four transmitting earth stations.



Each station transmits a preamble that contains synchronization and other data essential to the operation of the network before sending data. The earth stations transmission is followed by a guard time to avoid possible overlap of the following transmission. In GEO satellite systems, frame lengths of 125µs upto 2ms have been used, although 2ms has been used by stations using

Intelsat satellites. The transmitted bursts must contain synchronization and identification that help receiving earth stations to extract the required information without error. These goals are achieved by dividing TDMA transmissions into two parts : a preamble containing all the synchronization and identification data and a group of traffic bits. Synchronization of the TDMA network is achieved with the portion of the preamble transmitted by each earth station that contains carrier and bit clock synchronization waveforms. In some systems, a separate reference burst may be transmitted by one of the stations, designated as the master station. A reference burst is a preamble followed by no traffic bits. Traffic bits are the revenue producing portion of each frame, and the preamble and reference bursts represent overhead.

Figure shows a typical TDMA frame with 2ms duration used by some earth stations in TDMA.

CBTR	UW	TTY	SC	VOW	VOW	digital speech channels

CBTR – carrier and bit timing recovery

UW – Unique word

TTY – Teletype

SC – satellite channel

VOW – voice order wire

All of the blocks at the start of the frame, labeled CBTR through VOW, are preamble.

For the specific case of digital speech channels using serial transmission at a rate r_{sp}, the no. of speech channels n, that can be transmitted in a TDMA frame shared equally by N earth stations can be calculated from the duration of the frame, T_{frame} in seconds, the guard time t_g & preamble length t_{pre} in sec and the transmitted bit rate of the TDMA system R_b. The time T_d available in each station burst for transmission of data bits is

$$T_d = \frac{T_{frame} - N(t_g + t_{pre})}{N} \text{ seconds}$$

In is the total no. of bits, transmitted by each earth station is

$$C_b = \left[T_{frame} - N(t_g + t_{pre})\right] \frac{R_b}{T_{frame}}$$

Since each digital speech channel requires a continuous bit rate of r_{sp} bps, the no. of speech channels that can be carried by each earth station is given by n where

$$n = \left[T_{frame} - N\left(t_g + t_{pre}\right)\right] \frac{R_b}{T_{frame} \times \eta_{sp}}$$

Satellite switched TDMA

One advantage that TDMA has when used with a baseband processing transponder is satellite switched TDMA. Instead of using a single antenna beam to maintain continuous communication with its entire coverage zone, the satellite has a no. of narrow antenna beams that can be used sequentially to cover the zone. A narrow antenna beam has a higher gain than a broad beam, which ↑ the satellite EIRP and therefore ↑ the capacity of the downlink. Uplink signals received by the satellite are demodulated to recover the bit streams, which are structured as a seq of packets addressed to different receiving earth stations. The satellite creates TDMA frames of data that contain packets addressed to specific earth stations, and switches its transmit beam to the direction of the receiving earth station as the packets are transmitted. Note that control of the TDMA network timing could now be on board the satellite, rather than at a master earth station.

Onboard Processing:

The advantage of a bent pipe transponder is flexibility & disadvantage is that it is not well suited to uplinks from small earth stations, especially uplinks operating in Ka band. Consider a link between a small transmitting earth station and a large hub station via a bent pipe GEO satellite transponder. There will usually be a small rain fade margin on the uplink from the transmitting station because of its low EIRP. when rain affects the uplink, the C/N ratio in the transponder will fall. The overall C/N ratio in the hub station receiver cannot be greater than the C/N ratio in the transponder, so the bit error rate at the hub station will increase quickly as rain affects the uplink. The only available solution is to use forward error correction coding on the link, which lowers the data throughput but is actually needed for lessthan 5% of the time.

The problem of uplink attenuation in rain is most severe for 30/20 GHZ uplinks with small margins. Outages are likely to be frequent unless a large rain fade margin is included in the uplink power budget. Onboard processing & a baseband processing transponder can overcome this problem by separating the uplink and downlink signals and their C/N ratios. The baseband processing transponder can also have different modulation schemes on the uplink and downlink to improve spectral efficiency, and can dynamically apply forward error control to only those links affected by rain attenuation. All LEO satellites providing mobile telephone service & Ka band satellites providing internet access to individual users use onboard processing.

Demand access multiple access (DAMA):

Demand access can be used in any satellite communication link where traffic from an earth station is undefined. An example is an LEO satellite system providing links to mobile telephones. Telephone voice users communicate at random times, for periods ranging from <1 minute to several minutes. Demand access allows a satellite channel to be allocated to a user on demand, rather than continuously, which greatly increases the no.of simultaneous users who can served by the system.

Demand access systems require two different types of channel: a common signaling channel (CSC) & a communication channel.

A user ~~which~~ wishing to enter the communication network first calls the controlling earth station using the CSC, and the controller then allocates a pair of channels to that user. Packet transmission techniques are widely used in DA systems because of the need for addresses to determine the source and destination of signals. Bent pipe transponders are often used in demand access mode, allowing any configuration of FDMA channels to be adopted.

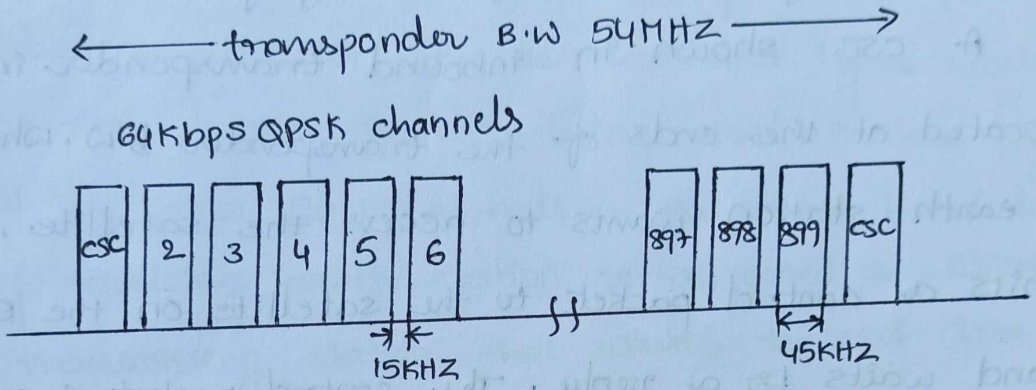


Figure shows a typical 54MHZ bandwidth Ku band transponder frequency plan for the inbound channels of a VSAT network using FDMA with SCPC & DA (FDMA-SCPC-DA) on the inbound link. The individual outbound RF channels are 45KHZ wide to accommodate the occupied bandwidth of 64Kbps bit streams transmitted using QPSK. A guard band of 15KHZ is allowed between each RF channel, so one RF channel requires a total bandwidth of 60KHZ. A 54MHZ bandwidth transponder can acommodate 900 of these 60KHz channels.

The outbound link of this particular VSAT network is a continuous TDM bit stream transmitted through a separate transponder. A and transponder is used to allow for the differences in transponder gain needed for the inbound and outbound channels of the VSAT system. In VSAT systems, the inbound and outbound channels are usually symmetric, offering the same data rate in opposite directions. Internet access systems are often assymmetric, because requests for information can be short but the resulting replies may be lengthy.

A CSC shown in inbound transponder in fig are located at the ends of the transponder B.W. When a VSAT earth station wants to access the satellite, it transmits a control packet to the satellite on the CSC freq and waits for a reply. The control packet is received by the hub earth station and decoded. The control packet contains the address of a terrestrial or satellite transponders destination for the call, DA, the address of the station requesting the connection, RA, any other relevant data and a CRC that is used in the receiver to check for errors in the packet. The control station records with both origination and destination station addresses and measures the duration of the connection in order to generate billing data.

Code division multiple access

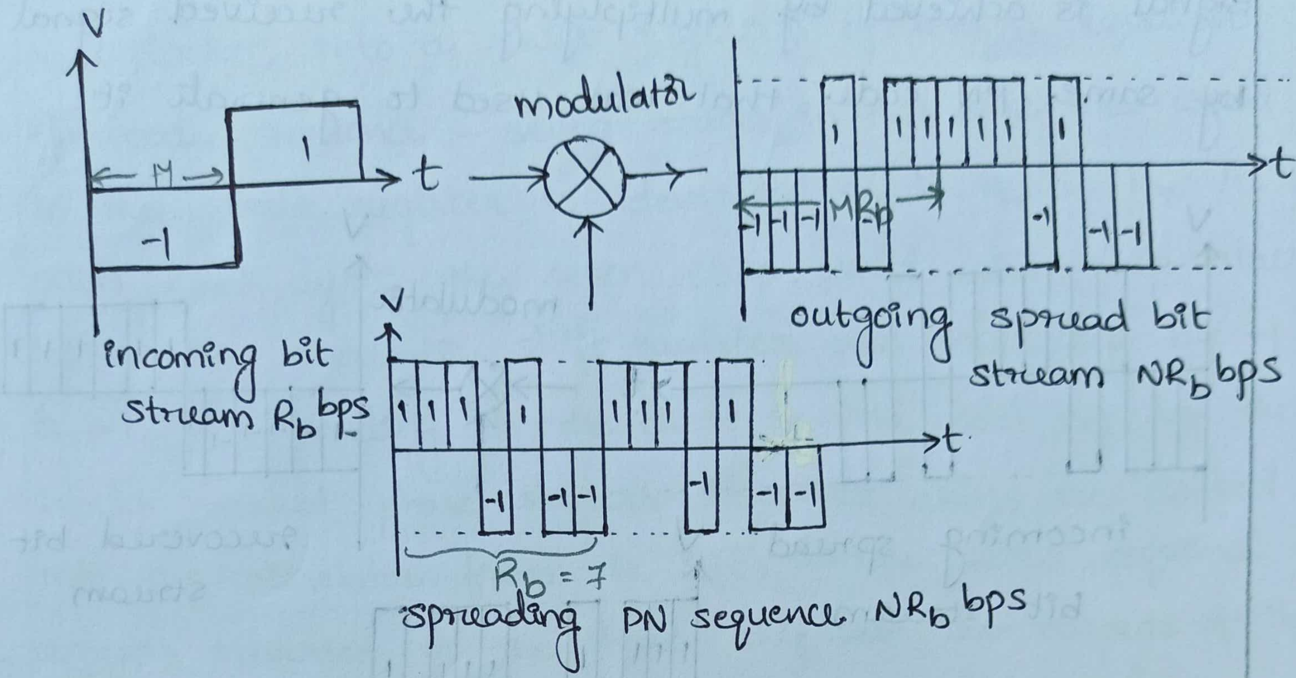
CDMA is a scheme in which a no.of users can occupy all of the transponder bandwidth all of the time. CDMA signals are encoded suchthat information from an individual transmitter can be recovered by a receiving station that knows the code being used, in the presence of all the other CDMA signals in the same bandwidth. Every receiving earth station is allocated a CDMA code any transmitting station that wants to send data to that earth station must use the correct code. CDMA codes are typically 16 bits to many thousands of bits in length, and the bits of a CDMA code are called chips to distinguish them from the message bits of a data transmission. The CDMA chip sequence modulates the data bits of the original message, and the chip rate is always much greater than the data rate. This greatly increases the speed of the digital transmission, widening its spectrum in propotition to the length of the chip sequence. As a result, CDMA is also known as spread spectrum. Direct sequence spread spectrum (DSSS) is the only type currently used in satellite communication; frequency hopping spread spectrum (FHSS) is used in the bluetooth system for multiple access in short range local area wireless networks.

CDMA has become popular in cellular telephone systems where it is used to enhance cell capacity. However it has not been widely adopted by satellite comm. systems because it usually proves to be less efficient, in terms of capacity than FDMA & TDMA. The GPS navigation systems uses DS-SS CDMA for the transmission of signals that permit precise location of a receiver in three dimensions.

Spread spectrum transmission and reception :

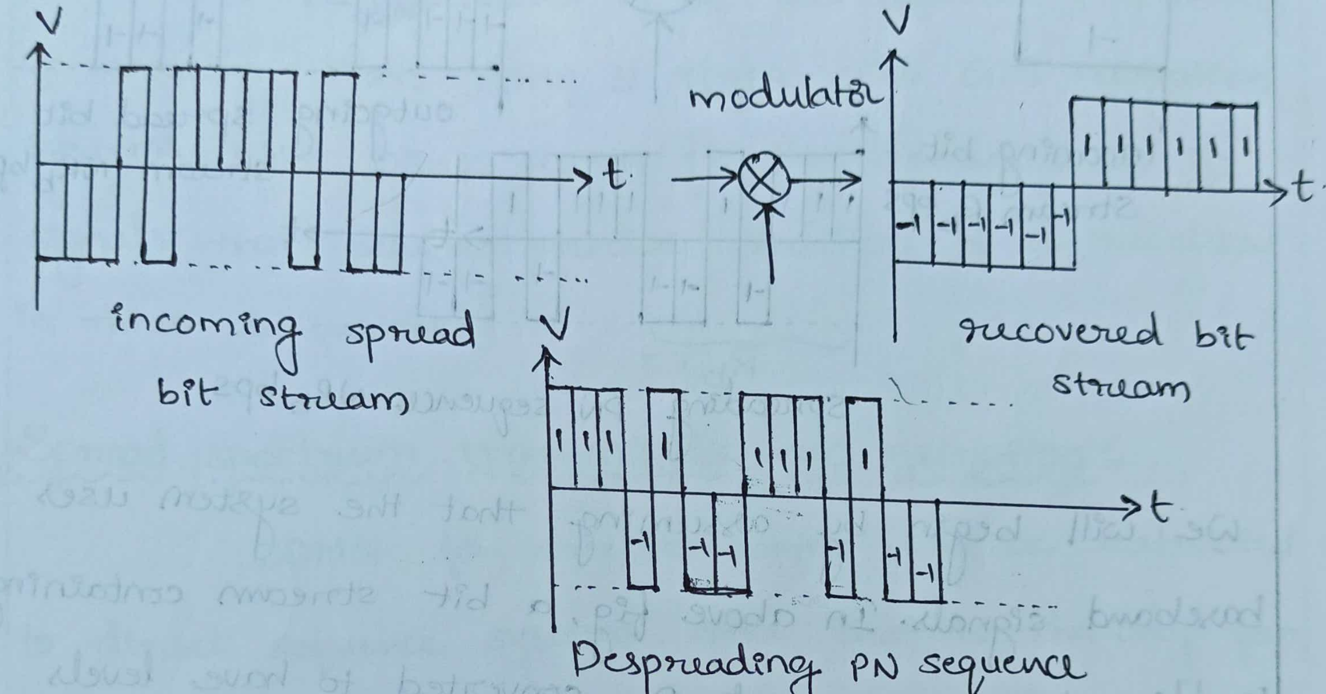
CDMA for satellite comm. will be restricted to direct sequence systems, since that is the only form of spread spectrum that has been used by commercial satellite systems to date. The spreading codes used in DS-SS CDMA systems are designed to have good autocorrelation properties and low cross correlation. Various codes have been developed specifically for this purpose, such as Gold and Kasami codes.

The DSSS codes will all be treated as pseudo-noise (PN) sequences in this discussions. Pesudo noise refers to the ~~sprectru~~ spectrum of code, which appears to be a random sequence of bits (or chips) with a flat, noiselike spectrum. The generation of a DSSS signal is illustrated in below fig.

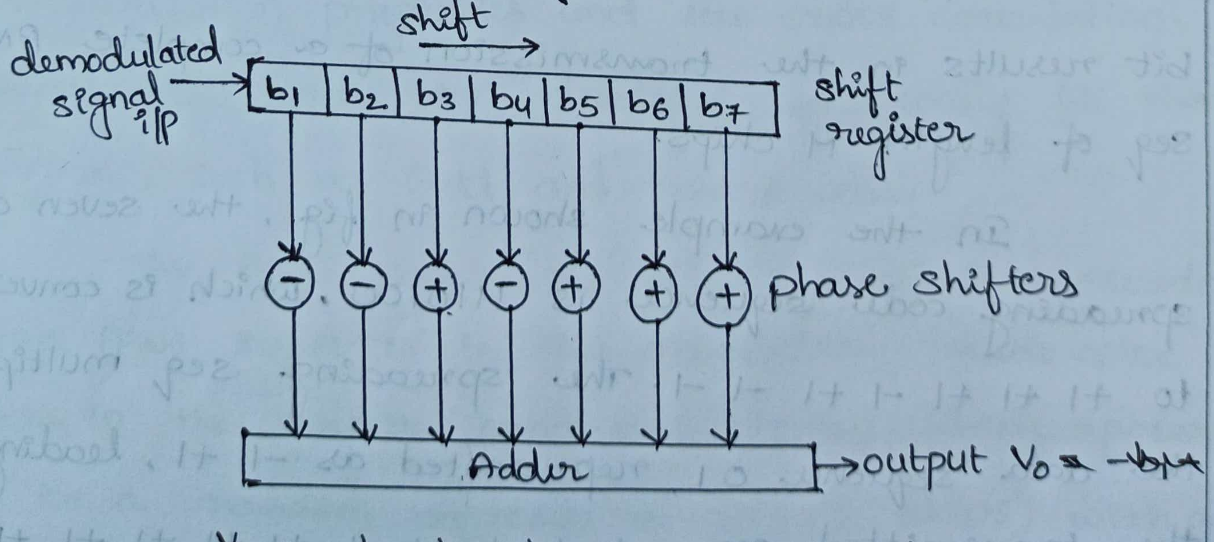


We will begin by assuming that the system uses baseband signals. In above fig, a bit stream containing traffic data at a rate R_b, converted to have levels of +1 and -1V corresponding to the logic states 1 and 0, is multiplied by a PN sequence, also with levels +1 and -1V at a rate $M \times R_b$ chips/second. Each data bit results in the transmission of a complete PN seq of length M chips.

In the example shown in fig, the seven chip spreading code sequence is 1110100, which is converted to +1 +1 +1 -1 +1 -1 -1. The spreading seq multiplies the data sequence 01, represented as -1 +1, leading to the transmitted seq -1 -1 -1 +1 -1 +1 +1 +1 +1 +1 -1 +1 -1 -1 shown at the right in above fig. Recovery of the original data stream of bits from the DSSS signal is achieved by multiplying the received signal by same PN code that was used to generate it



At a CDMA receiver which knows the seven bit code, there will be a correlator that has the code stored as multiplier settings.



$$V_0 = -b_1 - b_2 + b_3 - b_4 + b_5 + b_6 + b_7$$

Figure illustrates the correlation process. Received chips are clocked into a shift register of length equal to the code sequence - seven stages in this case. The word in the shift register is identified as $b_1, b_2, \ldots b_7$. At each clock cycle the seven chip word with chip values b_i in the correlator shift register are multiplied by +1 or -1, corresponding to the chips in the code seq, by the blocks marked phase shifters. Received chips are clocked into the correlator from the left, so the code sequence appears reversed in the phase shifters. The outputs of the phase shifters are added to give the output word V_0.

Satellite navigation and global positioning system

The global positioning satellite system has revolutionized navigation and position location. It is now the primary means of navigation for most ships and aircraft and is widely used in surveying and many other applications. The GPS system, originally called NAVSTAR, was developed as a military navigation system for guiding missiles, ships and aircraft to their targets. GPS satellites transmit L-band signals that are modulated by several codes (coarse acquisition code, P code) P→precise

The GPS system has been successful because it provides a direct readout of the present position of a GPS receiver with a typical accuracy of 30m. There are other position location systems, such as LORAN (long range navigation) → less accurate than GPS, less reliability than GPS.

The GPS space segment consists of 24 satellites in medium earth orbit (MEO) at a nominal altitude of 20,200Km with an orbital inclination of 55°. The satellites are clustered in groups of four, called constellations, with each constellation separated by 60° in longitude. The orbital period is approximately one half a sidereal day (11H - 58min) so the same satellites appear in the same position in the sky twice each day. The orbits of the 24GPS satellites ensure that at any time, anywhere in the world, a GPS receiver can pick up signals from atleast four satellites.

The position of a GPS receiver is found by trilateration, which is one of the simplest and most accurate methods of locating an unknown position. In trilateration, the distance of the unknown point from three known points is measured. The intersection of the arcs corresponding to three distances defines the unknown point relative to the known points, since three measurements can be used to solve three equations to give the latitude, longitude and elevation of the receiver. The distance between a transmitter and a receiver can be found by measuring the time it takes for a pulse of RF energy to travel between the two. Time can be measured electronically more accurately than any other parameter by the use of atomic clocks. Each satellite carries several high accuracy atomic clocks and radiates a sequence of bits that starts at a precisely known time. A Gps receiver contains a clock that is synchronized inturn to the clock on each satellite that it is receiving.

GPS satellites transmit two signals at different frequencies, known as L_1 and L_2. The L2 signal is modulated with a 10.23Mbps pseudorandom (PN) bit sequence called the P code ie used by military positioning systems. the P code is transmitted in an encrypted form known as the Y code, which restricts the use of Pcode to authorized users.

The L1 freq carrier is modulated by a 1.023Mbps PN sequence called the (coarse acquisition) C/A code that is available for public use, and also carries the pcode as a quadrature modulation. the higher bit rate of the P code provides better measurement accuracy than the 1.023Mbps C/A code. The GPS system provides two categories of service.

PPS - precise positioning service receiver

SPS - standard positiong service

The precise positioning receiver track both P code and C/A code on L1 and L2 frequencies. The PPS receivers is used mainly by military users. standard positioning service receivers track the C/A code on L1. this is the service used by general public. the P(Y) and C/A codes transmitted by each satellite create direct sequence spread spectrum signals. Both the C/A codes and the pcodes are publicly available, but the P code cannot be recoverd in a GPS receiver without a knowledge of the Y code decryption algorithm.

USA - GPS, china - Beidou

Russia - GLONASS (Global navigation satellite system)

European union - Galileo

Indian Regional navigation satellite system (IRNSS)

NAVIC (Navigation with Indian constellation)

Radio and satellite navigation:

Prior to the development of radio, navigation was by compass and landmarks on land, and by the sun and stars at sea. These are less accurate. Pilots of light aircraft, relying solely on a map and landmarks, would get lost and runout of fuel before they found somewhere to land. With a GPS receiver and a map, it is impossible to get lost. Gps receivers are very popular with airplane pilots, owners of sea going boats and wilderness hikers.

The development of aircraft that could fly above the clouds and particularly the building of large no.of bomber aircraft in the 1930's, made radio navigation essential. During WW-I & WW-II, placed high reliance on the ability of bomber aircraft to win a war by destroying the weapon manufacturing capability of the enemy. Bomber aircraft, ICBMs (Inter Continental Ballistic Missiles) and cruise missiles must find their targets, so accurate navigation is an essential part of each of these weapon systems. This demand for accurate targeting of airborne weapons led to the development of GPS.

commercial aircraft fly on federal airways using VOR (VHF omni range) beacons. The airways are 8 miles wide to allow for the angular accuracy of VOR measurements, which is better than 4°. GPS will eventually replace VOR navigation, allowing aircraft to fly directly from point of origin to destination, but the system of VOR ~~beao~~ beacons in the U.S is likely to remain for many years as a backup to GPS.

GPS can provide a single navigation system with better accuracy and reliability than all earlier radio navigation aids. It can provide navigation of aircraft directly between airports, instead of indirectly via airways, while providing absolute position readout of latitude and longitude. Differential GPS can be used instead of ILS to provide the required straight line in the sky for an instrument approach to a runway, and can be linked to an autopilot to provide automatic landing of aircraft in zero visibility conditions. ships can safely navigate and dock in treacherous waters in bad weather by using differential GPS.

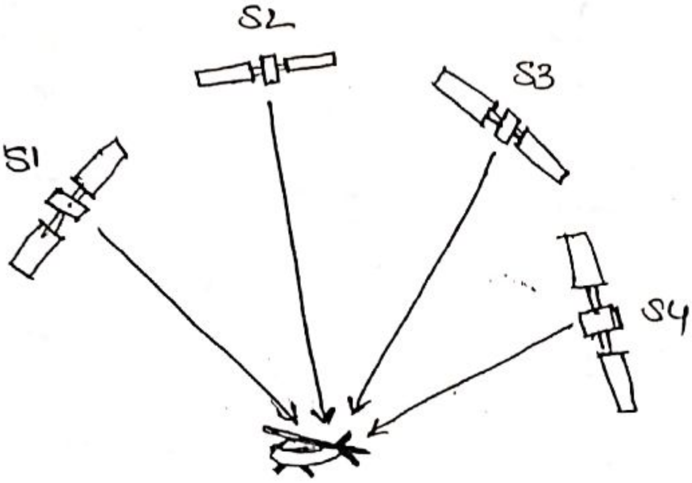
Gps was preceded by an earlier satellite navigation system called Transit, built for the U.S. Navy for ship navigation, which achieved much lower accuracy and became obsolete when GPS was introduced.

ILS - Instrument landing system

A similar system called SARSAT, for research and rescue satellite, is used to find emergency locator transmitters (ELTs) on aircraft that have crashed.

GPS position location principles

The basic requirement of a satellite navigation system like GPS is that there must be four satellites transmitting suitably coded signals from known positions. Three satellites are required to provide the three distance measurements, and the fourth to remove receiver clock error. Figure shows the general arrangement of position location with GPS.



The three satellites provide distance information when the GPS receiver makes three measurements of range R_i, from the receiver to three known points. Each distance R_i can be thought of as the radius of a sphere with a GPS satellite at its center. the receiver lies at the intersections of three such spheres, with a satellite at the center of each sphere. Locally, at the receiver, the spheres will appear to be planes since the radii of the spheres are very large. A basic principle of geometry is that the intersection of three planes completely defines a point. Thus three satellites, through measurement of their distances to the receiver, define the receiver location close to the earth's surface.

Although the principles by which GPS locates a receiver are very simple, requiring only the accurate measurement of three ranges to three satellites, implementing the measurement with the required accuracy is quite complex. Range is calculated from the time delay incurred by the satellite signal in traveling from the satellite to the GPS receiver, using the known velocity of EM waves in freespace. To measure the time delay, we must know the precise instant at which the signal was transmitted, and we must have a clock in the receiver that is synchronized to the clock on the satellite.

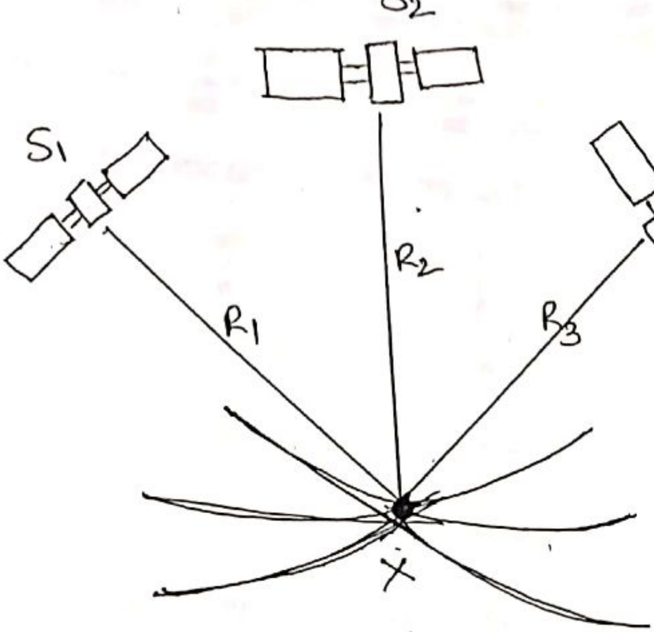
GPS satellites each carry four atomic clocks which are calibrated against time standards in the GPS control stations around the world. The result is GPS time, a time standard that is available in every GPS satellite. The accuracy of an atomic ~~clock~~ clock is typically 1 part in 10^{11}. However it is too expensive to include an atomic clock in most GPS receivers, so a standard crystal oscillator with an accuracy of 1 in 10^5 or 1 in 10^6 is used instead. The receiver clock is allowed to have an offset relative to the GPS satellite clocks, so when a time delay measurement is made, the measurement will have an error caused by the clock offset. For example, suppose the receiver clock has an offset of 10ms relative to GPS time. All distance measurements will then have an error of 3000km. Clearly, we must have a way to remove the time error from the receiver clock before we can make accurate position measurements. C/A code receivers can synchronize their internal clocks to GPS time with in 170ns, corresponding to a distance measurement uncertainty of 50m. Repeated measurements and integration improve the position location error to well below 50m.

It is easy to remove the clock error, and this removal is one of the strength of GPS. All that is needed is a time measurement from a fourth satellite. We need three time measurements to define the location of the receiver in the three unknown coordinates x, y, z. When we add a fourth time measurement we can solve the basic position location equations for a fourth unknown, the receiver clock offset τ. Four unknowns are x, y, z and τ

Position location in GPS

First, we will define the coordinates of the GPS receiver and the GPS satellites in a rectangular coordinate system with its origin at the center of the earth. this is called the earth centered earth fixed (ECEF) coordinate system, and is part of the WGS-84 (World Geodetic System) description of the earth. WGS-84 is an internationally agreed description of the earth's shape and parameters to calculate the orbits of the GPS satellites with the accuracy required for precise measurement of the range to the satellites. The z-axis of the coordinate system is directed through the earth's north pole and the x- and y-axes are in the equatorial plane. The x-axis passes through the Greenwich meridian, and the y-axis passes through the 90° east meridian. The ECEF coordinate system rotates with the earth. The receiver coordinates are (U_x, U_y, U_z) and the four satellites have coordinates (X_i, Y_i, Z_i) where $i = 1,2,3,4$ There may be more than 4 satellite signals available, but we use only four signals in a position calculation.

The measured distance to satellite i is called a pseudorange PR_i, because it uses the internal clock of the receiver to make a timing measurement that includes errors caused by receiver clock offset. The geometry of a GPS measurement is illustrated in below fig.



Pseudorange, denoted as PR_i, is measured from the propagation time delay T_i between the satellite (number i) and the GPS receiver, assuming that EM waves travel with velocity c.

$$PR_i = T_i \times C$$

The distance R between two points A and B in a rectangular coordinate system is given by

$$R^2 = (x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2$$

The equations which relate pseudorange to time delay are called ranging equations.

$$(X_1 - U_x)^2 + (Y_1 - U_y)^2 + (Z_1 - U_z)^2 = (PR_1 - \tau C)^2$$

$$(X_2 - U_x)^2 + (Y_2 - U_y)^2 + (Z_2 - U_z)^2 = (PR_2 - \tau C)^2$$

$$(X_3 - U_x)^2 + (Y_3 - U_y)^2 + (Z_3 - U_z)^2 = (PR_3 - \tau C)^2$$

$$(X_4 - U_x)^2 + (Y_4 - U_y)^2 + (Z_4 - U_z)^2 = (PR_4 - \tau C)^2$$

where τ = receiver clock error (offset or bias)

The position of the satellite at the instant it sent the timing signal (which is actually the start of a long seq of bits) is obtained from ephemeris data transmitted along with the timing signals. The receiver calculates the coordinates of the satellite relative to the center of the earth (X_i, Y_i, Z_i) and then solves the four ranging eqns for the four unknowns using standard numerical techniques for the solution of nonlinear eqns.

The four unknowns are the location of the GPS receiver (U_x, U_y, U_z) relative to the center of the earth & the clock offset - τ - called clock bias in GPS technology.

The receiver position is then referenced to the surface of the earth, and can be displayed in latitude, longitude and elevation. Typical accuracy for a low cost GPS receiver using the GPS C/A code is 30m defined as a 2DRMS error. DRMS means distance root mean square error of the measured position relative to the true position of the receiver.

The U.S dept of defense has the ability to degrade the position measurement accuracy of C/A code receivers by applying selective availability (SA). SA exists to allow the accuracy of C/A code receivers to be degraded in the event of a national emergency affecting the U.S. With SA off, the accuracy of GPS position measurements with the C/A code ↑ dramatically.

Selective availability and atmospheric propagation effects all cause errors in the timing measurements made by a GPS receiver, leading to position location errors. The errors can be largely removed if a no. of GPS reference stations are built at precisely known locations. The stations observe the GPS signals and compute the current error in position as calculated from GPS data. This information can then be broadcast to all GPS users as a set of corrections to be applied to GPS measurements. The system is called a wide area augmentation system (WAAS). Using WAAS accuracies of a few meters can be obtained with C/A code receivers. In the event of a national emergency, WAAS would be switched off to prevent enemies using GPS for accurate targeting of weapons.

Similarly, a single reference station at a known location can determine the local measurement error in GPS and broadcast this information to GPS users so that greater accuracy can be obtained with a C/A code receiver. This is one form of differential GPS (DGPS). With DGPS the receiver computes its position relative to the reference station rather than in latitude and longitude

GPS time

The clock bias value γ which is found as part of the the position calculation process can be added to the GPS receiver clock time to yield a time measurement i.e synchronized to the GPS time standard. Crystal oscillator used in the GPS receiver is highly stable over a period of a few seconds, but will have a freq which changes with temperature and with time. Temp changes cause the quartz crystal i.e the freq determining element of a crystal oscillator to expand or contract and this changes the oscillator freq.

Crystals also age, which causes the freq to change with time. The changes are very small, but sufficient to cause errors in the clock time at the receiver when the clock is not synchronized to a satellite. Calculating the clock bias by solving range eqns allows the receiver clock time to be updated every second or two so that the GPS receiver time readout is identical to GPS time.

The time standard on board each GPS satellite consists of two cesium clocks plus two rubidium clocks (atomic clocks). An atomic clock uses the fundamental resonance of the cesium or rubidium molecule as a freq reference to lock a crystal oscillator.

GPS receivers and codes

GPS satellites transmit using pseudorandom sequence (PN) codes. All satellites transmit a c/A code at the same carrier freq 1575.42MHZ (1.5GHZ) called L1 using BPSK modulation. The L1 freq is 154 times the master clock freq of 10.23MHZ. The c/A code has a clock rate of 1.023MHZ and the c/A code seq has 1023bits, so the PN seq lasts exactly 1ms.

The p code is transmitted using BPSK modulation at the L2 carrier freq of 1227.6MHZ (120 X 10.23MHZ) and is also transmitted with BPSK modulation on the L1 carrier freq, in phase quadrature with the C/A code BPSK modulation.

The C/A and P code transmissions from all GPS satellites are overlaid in the L1 and L2 freq bands, making GPS a direct sequence spread spectrum (DSSS) system. The receiver separates signals from individual GPS satellites using knowledge of the unique C/A code that is allocated to each satellite

The C/A code

The C/A codes transmitted by GPS satellites are all 1023 bit gold codes. GPS C/A gold codes are formed from two 1023 bit m-sequences, called G1 and G2. An m-sequence is a maximum length pseudorandom (PN) sequence, which is easy to generate with a shift register and feedback taps. A shift register with n stages can generate a PN sequence 2^n-1 bits in length. The PN seq G1 and G2 are both generated by 10 bit shift registers and are therefore both 1023 bits long. The clock rate for the C/A code is 1.023MHZ. so each seq lasts 1ms. Fig shows a generator diagram for the C/A code.

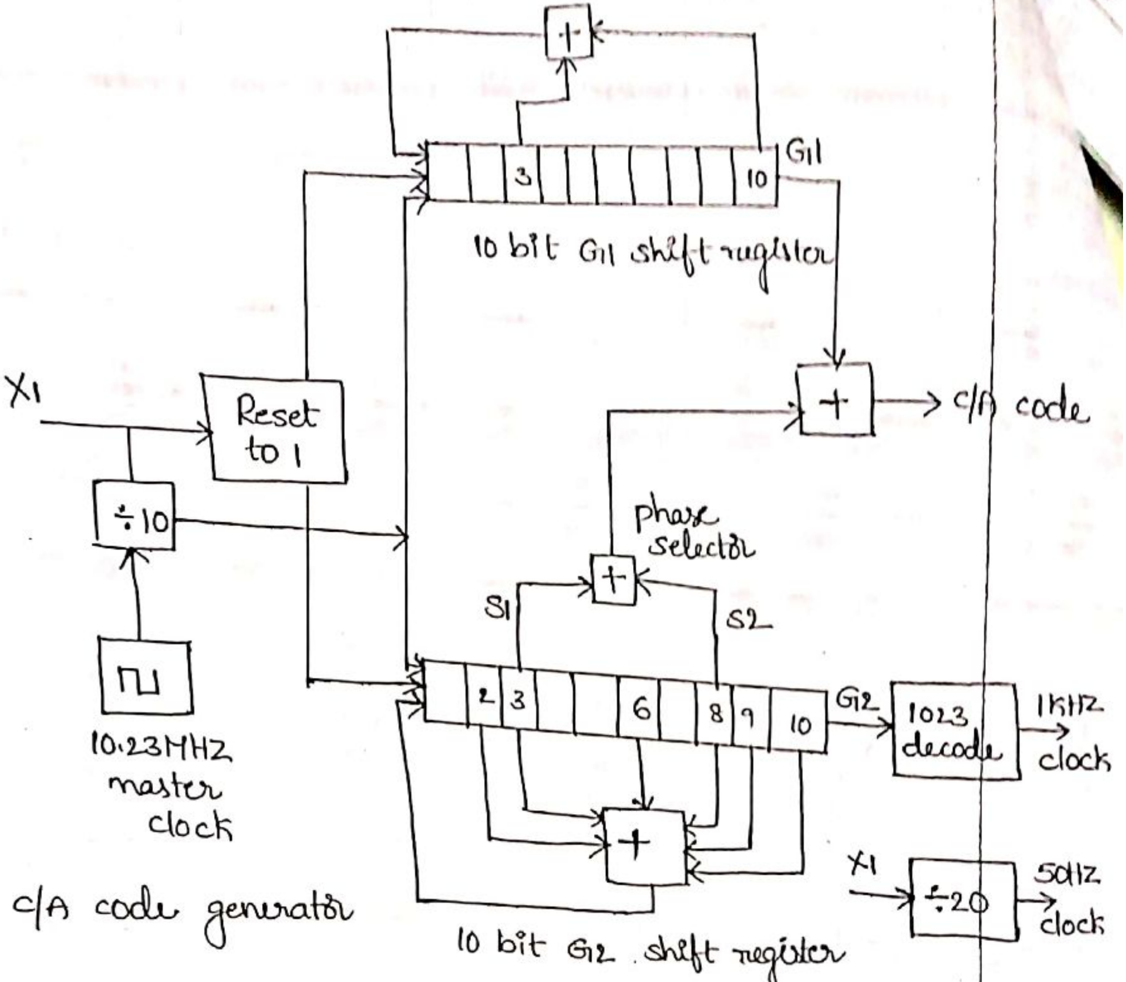


fig: c/A code generator

The c/A code for a particular satellite is created with an algorithm that includes the identification number of the GPS satellite, thus, creating a unique code for each satellite. The satellite with ID number i has a c/A code seq. $C_i(t)$.

$$C_i(t) = G_1(t) \times G_2(t+10iT_c)$$

where T_c = clock period for the c/A code.

A total of 100 gold sequences can be created using the algorithm in above eqn, but not all the sequences have sufficiently low cross correlation properties, only

37 are actually used in the GPS system. Low cross correlation of the sequences is a requirement because the GPS receiver can pick up signals from as many as 12 satellites at the same time. A correlator in the receiver looks for one of the sequences and must reject all other sequences that are present.

Figure shows a simplified block diagram of a C/A code GPS receiver.

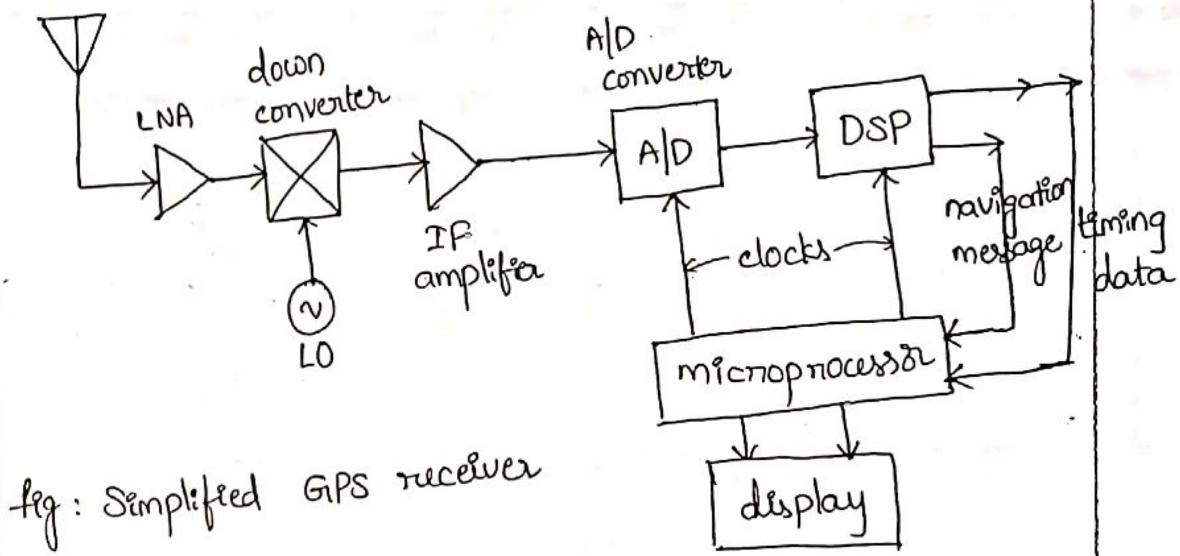


fig: Simplified GPS receiver

The antenna is typically a circularly polarized path antenna with an LNA mounted on the PCB. A super heterodyne receiver is used to generate an IF signal in a B.W of about 2MHz, which is sampled using I and Q sampling techniques and processed digitally. the digital portion of the receiver includes a C/A code generator, a correlator and a microprocessor that makes the timing measurements and calculates the receiver's position.

Satellite signal acquisition

The GPS receiver must find the starting time of the unique C/A code for each of four satellites. This is done by correlating the received signal with stored C/A codes, as in any direct sequence spread spectrum. Usually the receiver will automatically select the four strongest signals and correlate to those. If the geometry of the strongest satellites is poor i.e the satellites are close together and have pseudo ranges that are nearly equal, the receiver may also use several weaker signals. If the receiver is making a cold start, with no information about the current position of GPS satellites or its own location, it must search all 37 possible C/A codes until it can correlate with one. Once correlation is obtained, the data stream (navigation message) from that satellite can be read by the receiver.

A direct sequence spread spectrum receiver locks to a given code by matching the locally generated code to the code received from the wanted satellite. Since the start time of the code transmitted by the satellite is not known when the receiver commences the locking process, an arbitary ~~pa~~ start point must be selected. The locally generated code is compared to the received code, bit by bit, through all 1022 bits of the sequence, until either lock is found, or the receiver concludes that this is not the correct code for the satellite signal it is receiving.

If the starting time for the locally generated code was not selected correctly, correlation will not be obtained immediately. The locally generated code is then moved forward one bit in time, and correlation is attempted again. The process is continued 1023 times until all possible starting times for the locally generated code have been tried. If the satellite with that particular C/A code is not visible, no correlation will occur and lock will not be achieved. It takes a min of 1s to search all 1023 bit positions of a 1023 bit C/A code, so in a typical case, it will take atleast 15s to acquire the first satellite.

Although it takes only 20s on average to lock to the C/A code of one satellite, the receiver must find the Doppler freq offset for atleast one satellite before correlation can occur. The receiver B.W is matched to the B.W of C/A code. Once any of the GPS satellites has been acquired, the navigation message provides sufficient information about the adjacent satellites for the remaining visible satellites to be acquired quickly. The receiver may need to search in doppler shift because the position of the receiver relative to the satellites is not known, but their C/A codes are.

The correlation process described above assumes that each satellite is acquired sequentially. More sophisticated receivers have parallel correlators which can search for and acquire satellites in parallel. 12 parallel correlators guarantee that all visible GPS satellites will be acquired, and startup time is much shorter than with sequential acquisition. Accuracy is also better with parallel processing of the signals.

Integrity monitoring of the GPS position measurement is possible by using a 5th satellite to recalculate the receiver position. With five satellite signals there are five possible ways to select four pseudoranges to use in the ranging equations, leading to five calculations of position. If there is disagreement between the results, one bad measurement can be eliminated. If more than one result disagrees with the others, the integrity of the measurements is compromised. GPS receivers used for navigation of aircrafts in instrument meteorological conditions (IMC) and for instrument landings are required to have integrity monitoring to guard against receiver or satellite failures and interference with or jamming of GPS signals.

The P code for the ith satellite is generated in a similar way to the C/A code. The algorithm is $P_i(t) = X_1(t) + X_2(t + iT_c)$

where T_c = period of the X_1 seq contains 1,53,45,000 bits and repeats every 1.5S.

The X2 sequence is 37 bits longer. The Pcode repeats after 266.4 days, but is changed every 7days for security reasons. The C/A code provides information to authorized users on the starting time of the P code this is contained in the navigation message as an encrypted handover word. If the current feedback tap settings for the P code generators are known, and the handover word is decrypted, the receiver can start the local X code generators close to the correct point in the P code sequence. This allows rapid acquisition of the P code and is the origin of the name coarse acquisition for the C/A code.

GPS navigation message

A Key feature of the GPS C/A code is the navigation message. the navigation message is sent at 50bps by bpsk modulation of the C/A & P codes. Effectively 20 C/A code sequences form one navigation message bit. The phase of the 20 sequences is inverted between the 1 and 0 bits of the message by mod-2 addition of the navigation message data to the C/A and P code sequences. the navigation signal is extracted by a 50bps BPSK demodulator that follows the C/A or P code correlator.

The complete navigation message is 1500 bits, sent as a 30S frame with 5 subframes. Some information is contained in a sequence of frames, and the complete data set requires 12.5 min for transmission. The most important elements are of the message are repeated in every frame. The subframes contain the satellite's clock time data, orbital ephemeris for the satellite and its neighbors and various correction factors.

The calculation of position in a GPS receiver requires very accurate knowledge of the location of the satellite at the time that the measurements of pseudoranges are made. If the pseudorange is measured to an accuracy of 2.4m, we must know the satellite position to an even greater accuracy, and that requires very accurate calculation of the GPS satellite orbits. The GPS system uses modified WGS-84 data to define the earth's radius, Kepler's constant, and the earth's rotational rate. Data on the speed of EM waves is taken from the International Astronomical Union. All of these parameters and corrections are stored in every GPS receiver, and used in calculating position.

Header – Telemetry message: health of satellite, handover word

Subframe 1 – Satellite clock correction data, Age of transmitted data

Subframe 2 and 3 – Ephemeris for this satellite

Subframe 4 – Almanac data for satellites 25 and higher. Ionospheric model data

Subframe 5 – Almanac data for satellites 1–24. Health data for satellite 1–24.

Tab: GPS navigation message, subframe details.

GPS signal levels.

GPS receiver antennas have low gain because they must be omnidirectional. we will assume a worst case gain of G=odb, corresponding to an isotropic antenna In practice, G>odb in many directions, but may fall to odb in some directions. Typical GPS antennas are circularly polarized patches or quadrafilar helices that have carefully shaped patterns that cut off quickly below 10° elevation to minimize noise pickup from the ground. GPS satellites have an array of helical antennas that provide gain toward the earth, and low transmitters, leading to EIRP values in the range 19 to 27dbW. The C/A code transmitted by the satellite is a DSSS signal, so the C/N ratio in the C/A code's RF bandwidth will be lessthan odb. The low C/N ratio of the spread spectrum signal is converted to a usable S/N by correlation of the code sequences, which adds a despreading gain to the C/N ratio.

The GPS receiver can pick up signals from upto 10 satellites at the same time. The RF energy from the satellite spread spectrum transmissions adds to the noise in the receiver as an interference term I. For simplicity we will assume that there are 10GPS satellites visible, that there are 9 interfering satellites generating random signals (noise) out of which the receiver must extract the 10th signal, and that all the received signals are of equal strength. Nine interfering GPS satellites represents a worst case, in practice the no.of visible satellites varies between four and ten, and the signal strengths also vary depending on the elevation angle of satellite and the antenna pattern at the receiver. GPS receivers automatically select the strongest signals for processing, but if the sky is partially blocked by obstructions, a weak signal may have to be used.

The interference from nine C/A code spread spectrum signals of equal power is given by the sum of the received power from each satellite

GPS receiver operation

A C/A code GPS receiver must be able to correlate signals from atleast four satellites, calculate time delays, read the navigation message, calculate the orbits of the GPS satellites, and calculate position from pseudoranges. The key to accurate position determination is accuracy in the timing of the arrival of the Gold code sequences from each satellite in view. All GPS receivers use a microprocessor to make the required calculations and to control the display of data.

Most C/A code GPS receivers use an IC chip set that contain 12 parallel correlators. this allows the receiver to process signals from upto 12 satellites at the same time, which helps keep all the signals synchronized. Some simpler receivers use a single correlator and process four satellite signals sequentially, with consequent lower accuracy. The received GPS signals are converted to a suitable IF freq, in the front end of the receiver, and then processed to recover the C/A code. We will start the analysis by considering the signal received from the satellite at the output of the IF stage of the receiver.

The IF signal in the GPS receiver will consist of the sum of a no. of (12) signals from visible GPS satellites. The IF signal from N GPS satellites in view is

$$s(t) = \sum_{i=1}^{N} \left\{ A_i\, C_i(t) \cdot D_i(t) \cdot \sin\left[(\omega_i + \omega_d)t - \phi_i(l_i) + \phi_i \right] \right\}$$

where A_i = amplitude of the received signal

$C_i(t)$ = Gold code modulation

$D_i(t)$ = navigation message modulation

ω_i = IF freq of the received carrier

ω_d = doppler shift of the received signal

$\phi_i(l_i)$ = phase shift along the path

ϕ_i = phase angle of the transmitted signal

The key to successful measurements in a GPS C/A code receiver is to generate a signal in the receiver that is identical to the signal received from satellite i, but without the navigation data that is modulated onto the transmitted signal.

Part of a typical receiver structure for the GPS C/A code is shown in fig. The ~~function~~ of delay lock loop has three paths : punctual
early (half chip ahead)
late (half chip behind)

The delay lock loop steers the chip clock so that the punctual output can be used to drive the C/A code generator. The C/A code chip rate is generated by the VCO. The incremental process of trial and error which eventually

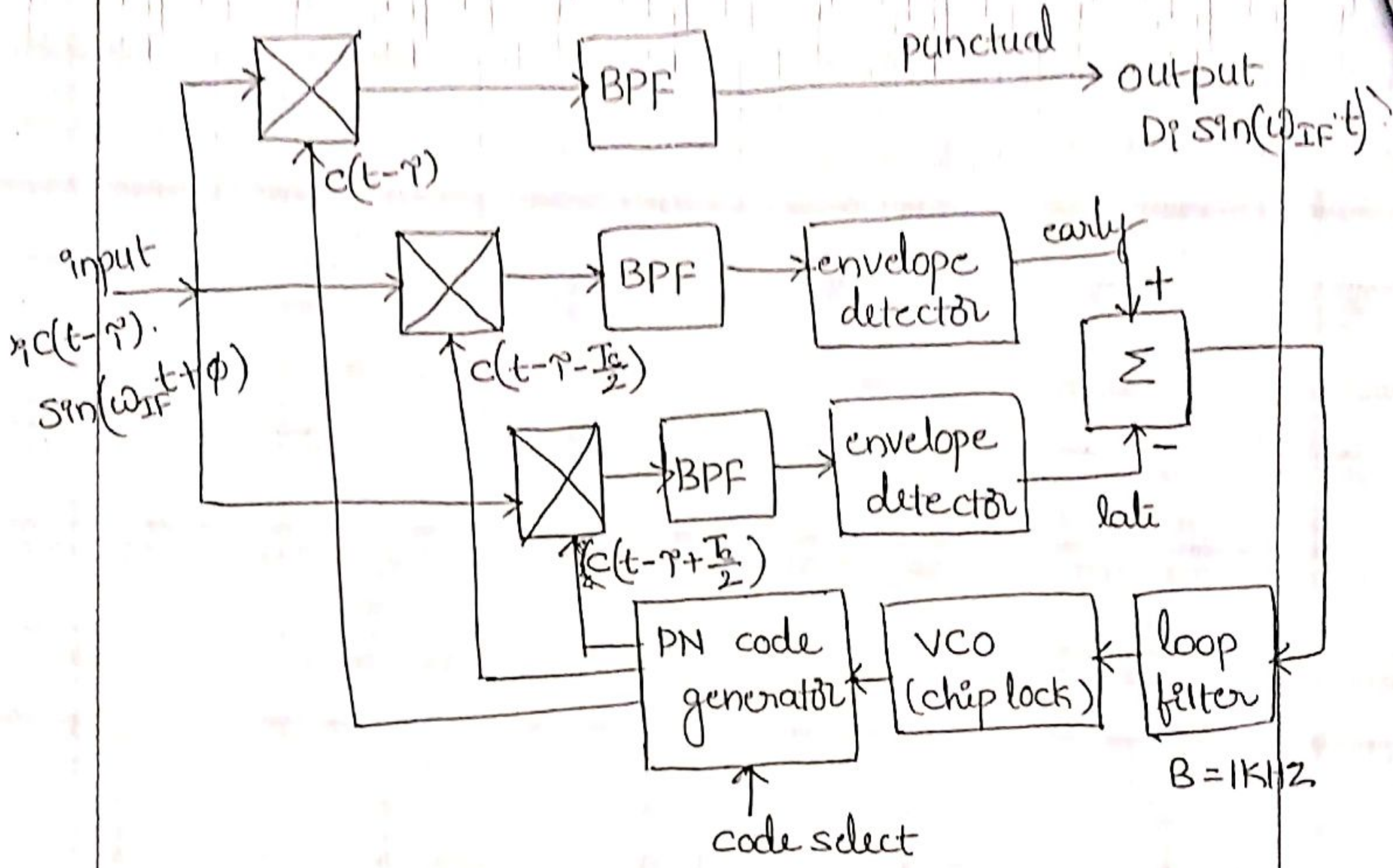


fig : Noncoherent code lock loop and navigation message recovery.

finds the correct sequence. and the early-late channels in the delay lock loop generate output signals which steer the phase of the VCO. so that navigation message is recovered correctly.

C/A code accuracy

The major sources of error in a GPS receiver that calculates its position are:

Satellite clock and ephemeris errors
Selective availability
Ionospheric delay
Tropospheric delay
Receiver noise
Multipath

The accuracy that can be achieved with a GPS C/A code receiver can be found by using a range error budget. Typical values of range error are given in table below.

Tab: Range error for C/A code measurements (m) meters

Satellite clock error	3.5
Ephemeris errors	4.3
Selective availability	32
Ionaspheric delay	6.4
Tropospheric delay	2
Receiver noise	2.4
Multipath	3
RMS range error	33.4 m with SA

The range error introduced by the troposphere can be partially removed by receiving identical signals at two different carrier frequencies. this technique is used by high precision p code receivers. The p code signal is transmitted on the L1 carrier at 1575.42MHZ, in phase quadrature with the C/A code signal. The P code is also transmitted on the L2 carrier at 1227.60MHZ.

The range error shown in above table is for one satellite-earth path; for the pseudorange that is calculated from the timing measurements using the receiver clock. However, four pseudorange measurements are needed to make a position determination. Thus the position location output of the GPS receiver combines four path errors, which are not necessarily equal because of the geometry of the satellites in the sky and the different signal strengths at the receiver input. Receiver position is calculated in (x, y, z) coordinates, and the errors in x, y, z depends on the elevation angle of satellites, the satellite geometry. The calculated position will have different levels of error in the x, y, z directions. To account for these differences several dilution of precision factors (DOP) are defined.

dilution of Precision ; HDOP, VDOP, GDOP

HDOP → Horizontal dilution of precision

VDOP → Vertical " "

GDOP → Geometric " "

PDOP → Position " "

TDOP → Time " "

HDOP is one of the most important DOP factors for most GPS users. It provides an error metric for the x and y directions, in the horizontal plane. In general, VDOP and GDOP are most likely to degrade the accuracy of GPS position measurements. VDOP accounts for loss of accuracy in the vertical direction caused by the angles at which the satellites being used for the position measurement are seen in the sky. VDOP is important in aircraft position measurements, where height above the ground is a critical factor, especially when landing. C/A code GPS receivers cannot guarantee sufficient vertical accuracy unless operated in a differential GPS mode.

The GPS satellites are configured in orbit to minimize the probability that a DOP can become large by arranging the orbits to provide clusters of four satellites with suitable spacings in the sky. Aircraft and ships at sea always have a clear view of the sky but automobiles do not. C/A code receivers may revert to two dimensional measurements (x and y) using three satellites when the sky is obstructed.

Differential GPS :

The accuracy of GPS satellites measurements can be increased considerably by using differential GPS (DGPS) techniques. A second, fixed, GPS receiver at a reference station is always required in a differential GPS system. A second GPS receiver at a known position continuously calculates its position using the GPS C/A code. The calculated (x, y, z) location is compared to the known location of the station and the differences in x, y, z are sent by a radio telemetry link to the first GPS receiver. This technique works well only if the two stations are close together and use the same four satellites for the position calculation.

In a more sophisticated form of differential GPS, the monitoring station at a known location measures the error in pseudorange to each satellite that is visible at its location, and telemeters the error values to users in that area. The most accurate forms of DGPS use the relative phase of the many signals in the GPS transmissions to increase the accuracy of the timing measurements. In principle, measurements which compare the phase angle of he received L1 carriers from several GPS satellites could therefore be used to detect receiver movements at the centimeter level. This is called differential phase or Kinematic DGPS.