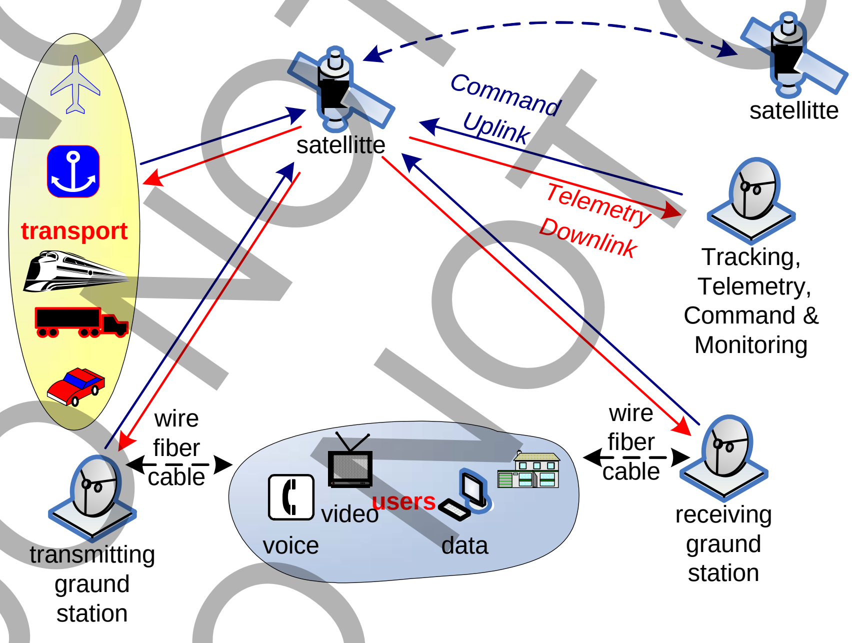
Lectures
on
Satellite Communications Systems

Assoc. Professor PhD Stefan Zhelev
Shumen University
Faculty of Technical Sciences



$$P_R = EIRP + G_R - L_{FS},\ \ dB$$

ISBN 978-954-577-637-3
електронно издание

Шумен, 2012

Contents

2.2. Frequency Division Multiple Access with PCM/SCPC/PSK application

3. Time Division Multiple Access
   3.1. Time Division Multiple Access with PCM/TDM/PSK application
   3.2. TDMA Frame Efficiency
   3.3. TDMA capacity
   3.4. Switching in satellite TDMA

4. Code Division Multiple Access
   4.1. Direct Sequence Spread Spectrum
   4.2. Frequency Hopping Spread Spectrum.

Problems

DO NOT COPY
DO NOT COPY

Satellite Communications Systems
Lecture 1.
Introduction to Satellite Communications

1. A little History of Satellite Communications
2. Satellite Communications Segments
3. Satellite Orbits
4. Frequency Band
4.1. Regulatory Process
4.2. The electromagnetic frequency spectrum
5. Signal processing elements in satellite communications

1. A little History of Satellite Communications

The unique feature of communications by satellites is there ability to provide distance insensitive point to multipoint communications. These links can be between fixed terminals, mobile terminals and both fixed and mobile terminals on lend, on the air, and at sea as it is shown in Figure 1.

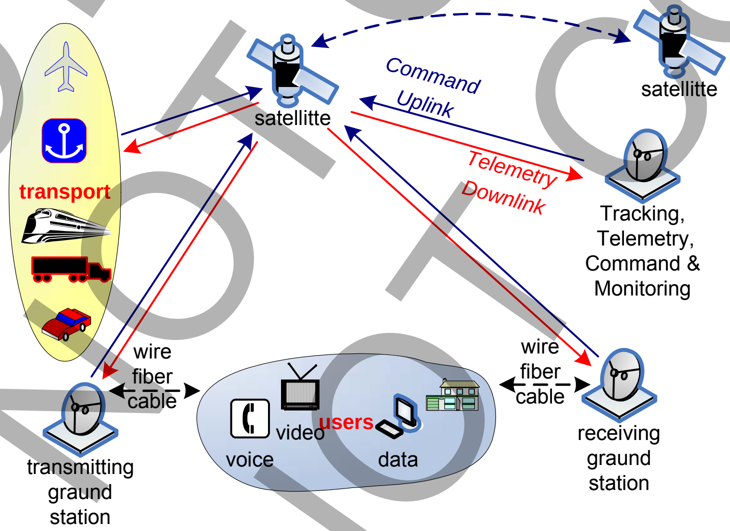


Figure 1. The idea of a communications satellite network

Communications by satellites offers a number of features that are not readily available with alternative modes of transmission, such as terrestrial microwave, cable, or fiber networks.

Some of the ***advantages*** of satellite communications are:

Diverse User Networks. The typical communications satellite can see large areas of the earth, and link together many users simultaneously. Satellites are particularly useful for accessing remote areas or communities not otherwise accessible by terrestrial means.

High Capacity. Satellite communications links operate with high carrier frequencies, with large information bandwidths. Capacities of typical communications satellites are very high (above $100 Mbps$), and can provide services for several hundred video channels or several tens of thousands of voice or data links.

Low Error Rates. Bit Error Rate (BER) is a measure of quality. Bit errors on a digital satellite link tend to be random, allowing statistical detection and error correction techniques to be used. Error rates can be routinely achieved efficiently and reliably with standard equipment.

Distance Independent Costs. The cost of satellite transmission is independent of the distance between the transmitting and receiving earth stations.

Fixed Broadcast Costs. The cost of satellite broadcast transmission, that is, transmission from one transmit ground terminal to a number of receiving ground terminals, is independent of the number of ground terminals receiving the transmission.

A little History of Satellite Communications

The first idea of communications by satellite was given by Arthur C. Clarke in his classic paper “Extraterrestrial Relays”, 1945[1].

Satellite communications began in October 1957 with the launch by the former USSR a small satellite called Sputnik 1 (4.10.1957), then 3.11.1957 was launched Sputnik 2 with Laika.

The **first communications by artificial satellite** was accomplished by SCORE (Signal Communicating by Orbiting Relay Equipment), launched by the Air Force, USA into a low (160 by 1280 km) orbit in December 1958. SCORE relayed a recorded voice message, from one earth station to another. SCORE broadcast a message from President Eisenhower to stations around the world. The maximum message length was four minutes, and the relay operated on a 150MHz uplink and 108MHz downlink. SCORE, powered by battery only, operated for 12 days before its battery failed, and decayed out of orbit 22 days later[2].

1960’s – First satellite communications:

- 1960 First passive communication satellite (Large balloons, ECHO satellites 1 and 2, launched by the National Aeronautics and Space Administration (NASA) in August 1960 and January 1964, respectively).
- 1962: First active communication satellite (The TELSTAR Satellites 1 and 2, launched into low orbits by NASA for AT&T/Bell Telephone Laboratories in July 1962 and May 1963, were the *first active wideband communications satellites*).
- 1963,1964: First satellite into geostationary (GEO) orbit (Syncom 1 failed, SYNCOM 2 and 3 were placed in orbit in July 1963 and July 1964, Syncom, with $7{,}4 GHz$ uplink and $1{,}8 GHz$ downlink frequencies, employed two $500 KHz$ channels for two-way narrowband communications, and one $5 MHz$ channel for one-way wideband transmission).
- 1964: International Telecomm. Satellite Organization (INTELSAT) created.
- 1965 First successful communications GEO (Early Bird / INTELSAT 1).

[1] A.C. Clarke, ‘Extraterrestrial Relays,’ *Wireless World*, Vol. 51, pp. 305 308, October 1945.
[2] M.I. Davis and G.N. Krassner, ‘SCORE First Communications Satellite,’ *Journal of American Rocket Society*, Vol. 4, May 1959.

1970's – GEO Applications Development, DBS:

- 1972 First domestic satellite system operational (ANIKA launched in November 1972 by NASA for Telsat Canada, was the *first domestic commercial communications satellite*. Two later ANIKAs were launched in April 1973 and May 1975. The satellites, built by HughesAircraft Company, operated at C-band and had 12 transponders, each 36MHz wide).
- 1975 First successful direct broadcast experiment (USA-India).
- 1977 A plan for direct broadcast sattellites (DBS) assigned by the ITU
- 1979 International Mobile Satellite Organization (Inmarsat) established.

1980's – GEO Applications Expanded, Mobile:

- 1981 First reusable launch vehicle flight.
- 1982 International maritime communications made operational.
- 1984 First direct-to-home broadcast system operational (Japan).
- 1987 Successful trials of land-mobile communications (Inmarsat).
- 1987 TVSAT: First DBS-satellite (Direct Broadcast Satellite, Television-broadcasts directly to home)
- 1989-90 Global mobile communication service extended to land mobile and aeronautical use (Inmarsat)

1990+'s NGSO applications development and GEO expansion

- 1990-95:

- Proposals of non-geostationary (NGSO) systems for mobile communications.

- Continuing growth of VSATs (Very Small Aperture Terminal) around the world.

- Spectrum allocation for non-GEO systems.

- Continuing growth of DBS. DirectTV created.

- 1997:

- Launch of first batch of LEO for hand-held terminals (Iridium).

- Voice-service portables and paging-service pocket size mobile terminals launched (Inmarsat).

- 1998-2000: Mobile LEO systems initiate service and fail afterwards (Iridium, Globalstar).

2. Satellite Communications Segments

The satellite communications have two areas or segments: the **space segment** and the **ground** (or **earth**) **segment**.

The space segment includes the satellite (or satellites) in orbit in the system, and the ground station that provides the operational control of the satellite(s) in orbit. The ground station is variously referred to as the *Tracking, Telemetry, Command (TT&C)* or the *Tracking, Telemetry, Command and Monitoring (TTC&M)* station. Two elements of the space segment of a communications satellite system are shown on Figure 2 - the satellite and TTC&M station.

A communications satellite is an orbiting artificial earth satellite. It receives a communications signal from a transmitting ground station, amplifies or processes it, and transmits it back to the ground stations for reception.

The TTC&M station provides essential spacecraft management and control functions to keep the satellite operating safely in orbit. The TTC&M links between the spacecraft and the ground are usually separate from the user communications links and links may operate in the same frequency bands or in other bands. The space segment equipment carried aboard the satellite has two subsystems: the *bus* and the *payload*.

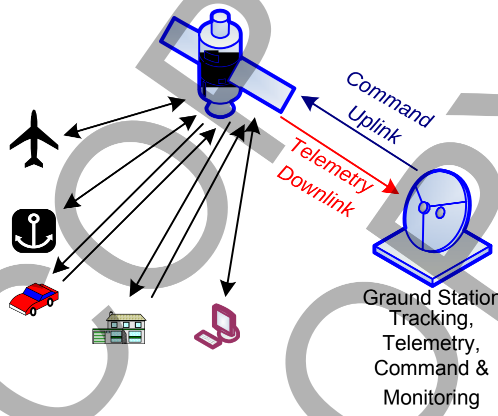


Figure 2. The space segment for a communications satellite network

The bus has the subsystems that support the satellite: the physical structure, power subsystem, attitude and orbital control subsystem, thermal control subsystem, and command and telemetry subsystem.

The payload on a satellite is the equipment that provides the services, and consists of the communications equipment that provides the relay link between the up- and downlinks from the ground. That equipment is called the ***transponder***. The antennas on the satellite and the transponder receives the uplink signal, amplifies (or processes the signal), and then reformats and transmits the signal back to the ground received station.

That portion of the link from the earth station to the satellite is called the ***uplink***, and another - from the satellite to the ground is called the ***downlink***.

The ground segment terminals consist of three basic types:

- fixed (in-place) terminals;
- transportable terminals;
- mobile terminals.

The TTC&M ground station are not included in the ground segment.

Fixed terminals may be providing different types of services, but they are designed to access the satellite while fixed in-place on the ground. Small terminals used in private networks (VSATs) are examples of fixed terminals, or terminals mounted on residence buildings used to receive broadcast satellite signals.

Transportable terminals are designed to be movable. Examples of the transportable terminal are satellite news gathering (SGN) trucks, which move to locations, stop in place, and then deploy an antenna to establish links to the satellite.

Mobile terminals are designed to communicate with the satellite while in motion. They are further defined as land mobile, aeronautical mobile, or maritime mobile, depending on their locations.

3. Satellite Orbits

There are four most commonly used orbits in satellite communications.

Geostationary (Geosynchronous) Earth Orbit (GEO)

The satellite remains fixed (or approximately fixed) over one point on the equator. The parameters for the evaluation of the GEO link are:

Range (distance) from the earth station (ES) to the satellite, in km

Geostationary Radius: $r_S = 42164{,}17 km$;

Geostationary Height (Altitude): $h_{GSO} = r_s - r_e = 35786{,}43 km$;

Equatorial Radius: $r_e = 6378{,}14 km$;

Advantages of Geostationary Earth Orbit:

- The period of revolution for the geostationary orbit is 23 h 56 min 4,091 s, so the visibility of the satellite is 24 h;
- A satellite in GSO sees about one-third of the earth's surface, so three GEO satellites, placed 120^0 apart in the equatorial plane, could provide global coverage;
- An antennas on the ground, once aimed at the satellite, need not continue to rotate.

Disadvantages of the GEO are:

- the gravity of the sun and moon disturb the orbit;
- the geostationary orbit's finite capacity and satellites using the same frequencies must be separated to prevent mutual interference;
- providing coverage of high latitudes (above 80^0) is generally not possible, so the polar aeries cannot be achieved with a geostationary constellation.

Low Earth Orbit (LEO)

LEO satellites move faster than the rotation of earth. The typically altitudes are from 160 to 2500 km. The satellite is not at a fixed location in the sky and the visibility of the satellite is $5 \div 15 \min$, so it needed many satellites to see the earth's surface. Some current LEO satellite networks operate with 12, 24, or 66 satellites to achieve the desired coverage.

Applications of LEO satellites are: Communications; Military surveillance; Weather; Atmospheric studies; Earth observation.

Medium Earth Orbit (MEO)

Satellites at Medium Earth Orbit (MEO) operate in the range between LEO and GSO, typically at altitudes of 10 000 to 15 000 km.

Applications: used for meteorological, remote sensing and position location applications (GPS).

High Elliptic Earth Orbit (HEO)

Satellites operating in high elliptical (high eccentricity) orbits (HEO) are used to provide coverage to high latitude areas not reachable by GSO, and those that require longer contact periods than available with LEO satellites. Russian use of Molnya and Tundra elliptical orbits for satellite television to the high northern latitude

Russian states is well known. These orbits have a perigee altitude of about 1000 km, and an apogee altitude of nearly 40 000 km.

Satellite orbits that are not synchronous, such as the LEO, MEO, or HEO, are often referred to as non-geosynchronous orbit (NGSO) satellites.

Orbital parameters are:

- *Apogee* – the point farthest from earth;
- *Perigee* – the point of closest approach to earth;
- *Line of Apsides* – the line joining the perigee and apogee through the center of the earth;
- *Ascending Node* – the point where the orbit crosses the equatorial plane, going from south to north;
- *Descending Node* – the point where the orbit crosses the equatorial plane, going from north to south;
- *Line of Nodes* – the line joining the ascending and descending nodes through the center of the earth;
- *Argument of Perigee* – the angle from ascending node to perigee, measured in the *orbital* plane;
- *Right Ascension of the Ascending Node*, – the angle measured eastward, in the equatorial plane, from the line to the first point of Aries (Y) to the ascending node.

Table 1. The characteristics of satellite orbits

	LEO	**MEO**	**GEO**	**HEO**
Altitude, km	500÷3 000	10 000÷14 000	35 786	500÷50 000
Period, h	1÷3	6÷8	23 h 56 min. 4,091s one sidereal day	3÷24
Delay, ms	6÷30	70÷120	240÷280	50÷320
Visibility	5÷15 min.	some hours	24 h	2÷12 h
Tracking, Telemetry, Command & Monitoring	complexity	less complexity	simply	complexity
Outlay for launch a satellite into orbit	low	high	high	high
Numbers of satellites to cover the earth	66	<24	3	
Applications	- Communications; - Military surveillance; - Weather; - Atmospheric studies; - Earth observation.	GPS	- telephony; - Broadcasting; - Point to multi-point communications; - Mobile services; - Weather observation.	

The parameters for the evaluation of the GEO link are:

- Range (d-distance) from the earth station (ES) to the satellite, in km;
- Azimuth angle (Θ) from the ES to the satellite, in degrees. The azimuth angle is the angle measured in the horizontal plane of the location between the direction of geographic north and the intersection of the plane containing the point considered, the satellite and the centre of the earth.
- Elevation angle (φ) from the ES to the satellite, in degrees. The elevation angle is the angle between the horizon at the point considered and the satellite, meas-

ured in the plane containing the point considered, the satellite and the centre of the earth.

Additional parameters are:

Equatorial Radius: $r_e = 6378{,}14 km$; Eccentricity of the earth: $e_e = 0{,}08182$; Differential longitude, $B = l_e - l_s$ - the difference between the earth station and satellite longitudes; l_e earth station longitude, in degrees; l_s satellite longitude, in degrees; L_E earth station latitude, in degrees; L_S satellite latitude, in degrees; H earth station altitude above sea level, in km.

Longitudes east of the Greenwich Meridian and latitudes north of the Equator are positive; longitudes west of the Greenwich Meridian and latitudes south of the Equator are negative.

The distance between earth station and satellite can be calculated as[3]:

$$d = \sqrt{R^2 + r_s^2 - 2R.r_s.\cos\psi_E.\cos B}\ ,\quad km \tag{1}$$

where $R = \sqrt{l^2 + z^2}$; $\quad l = \left(\dfrac{r_e}{\sqrt{1 - e_e^2.\sin^2(L_E)}} + H\right).\cos(L_E)$;

$$z = \left(\frac{r_e(1 - e_e^2)}{\sqrt{1 - e_e^2 \sin^2(L_E)}} + H\right).\sin(L_E); \qquad \psi_E = tg^{-1}\left(\frac{z}{l}\right).$$

Elevation angle to the satellite:

$$\varphi = \cos^{-1}\left(\frac{r_e + h}{d}\sqrt{1 - \cos^2(B).\cos^2(L_E)}\right). \tag{2}$$

Azimuth angle to the satellite:

The azimuth angle is determined from the intermediate angle A from one of four possible conditions, based on the relative location of the earth station and the subsatellite point on the earth's surface. The condition is determined by standing at the earth station (ES) and looking in the direction of the subsatellite point (SS)[4] (Figure 3).

$$A = \sin^{-1}\left(\frac{\sin(|B|)}{\sin\beta}\right), \tag{3}$$

where $\beta = \cos^{-1}[\cos(B).\cos(L_E)]$, $|B| = l_e - l_s$.

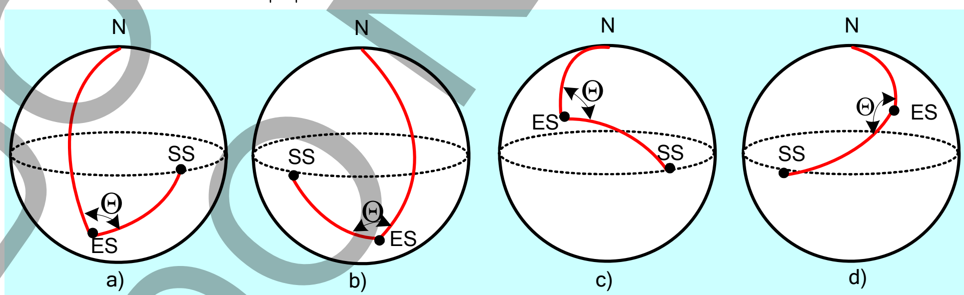


Figure 3. Determination of azimuth angle to the satellite

[3] Ippolito, Louis J., Satellite communications systems engineering: atmospheric effects, satellite link design, and system Performance, 2008 JohnWiley & Sons Ltd.

[4] Ippolito, Louis J., Satellite communications systems engineering: atmospheric effects, satellite link design, and system Performance, 2008 JohnWiley & Sons Ltd.

The resulting equation to determine the azimuth angle Θ for each of the four conditions is given in Table 2.

Table 2. Determination of azimuth angle from intermediate angle A

Condition SS point of ES	$\Theta =$	Figure
northeast	A	1a
northwest	$360^0 - \text{A}$	1b
southeast	$180^0 - \text{A}$	1c
southwest	$180^0 + \text{A}$	1d

More then four possible conditions, there are two additional cases:

- If the earth station is located at the same longitude as the subsatellite point, the azimuth angle will be 180^0 if the earth station is in the northern hemisphere and 0◦ if the earth station is in the southern hemisphere.

- If the earth station is located on the equator, the azimuth angle will be 90◦ if the earth station is west of the subsatellite point and 270^0 if the earth station is east of the subsatellite point.

Example 1. The earth station is located at Coventry, England.

Latitude: L_E =52°25'=52.42^0, N=+52.42; Longitude: $01^0 28' W = 1.47 W$, $l_e = -1.47$; Altitude: H=0km.

Satellite: Longitude: 19.2^0 E, $l_S = +19{,}2$.

Find the rang, azimuth angle (Θ) and elevation angle (φ).

1) Determine the differential longitude

$B = l_e - l_s = -1.47 - 19.2 = -20{,}67$;

2) Determine the earth radius at the earth station, R, for the calculation of the range

$$l = \left(\frac{r_e}{\sqrt{1 - e_e^2 . \sin^2(L_E)}} + H \right) . \cos(L_E) = \left(\frac{6378{,}14}{\sqrt{1 - 0{,}08182^2 . \sin^2(52{,}42)}} + 0 \right) . \cos(52{,}42) = 3897{,}8 km$$

$$z = \left(\frac{r_e(1 - e_e^2)}{\sqrt{1 - e_e^2 . \sin^2(L_E)}} + H \right) . \sin(L_E) = \left(\frac{6378{,}14.0{,}9933}{0{,}998} + 0 \right) . 0{,}7925 = 5030{,}87 \ \ km$$

$$\psi_E = tg^{-1}\left(\frac{z}{l} \right) = 52{,}23^0; \ R = \sqrt{l^2 + z^2} = 6364{,}15 km$$

3) Determine the range

$$d = \sqrt{R^2 + r_s{}^2 - 2 R . r_s . \cos\psi_E . \cos B} =$$

$$= \sqrt{6367{,}84^2 + 42164{,}17^2 - 2.6367{,}84.42164{,}17.0{,}6125.0{,}9356} = 38866{,}9 km$$

4) Determine the elevation angle

$$\varphi = \cos^{-1}\left(\frac{r_e + h}{d} \sqrt{1 - \cos^2(B) . \cos^2(L_E)} \right) = 27{,}01^0$$

5) Determine the azimuth angle

$$\beta = \cos^{-1}\left[\cos(B) . \cos(L_E) \right] = \cos^{-1}(0{,}9356.0{,}6099) = 55{,}21^0$$

$$\text{A} = \sin^{-1}\left(\frac{\sin(|B|)}{\sin\beta} \right) = \sin^{-1}\left(\frac{0{,}353}{0{,}821} \right) = 25{,}47^0$$

Since the subsatellite point SS is southeast of the earth station ES according Table 2: $\Theta = 180^0 - A = 154{,}53^0$.

4. Frequency Band

4.1. Regulatory Process

The satellite communications system parameters that are under the regulatory include:

- choice of radiating frequency;
- maximum allowable radiated power;
- orbit locations (slots) for GSO.

The allocation and regulation of the frequency spectrum is colled *spectrum* (or *frequency*) *management*.

The Europe countries have active organizations involved with spectrum management. They are responsible for the development of satellite systems or the provision of satellite based services. Besides national organizations there is international management by the International Telecommunications Union (ITU), with headquarters in Geneva, Switzerland. The ITU was formed in 1932 from the International Telegraph Union, created in 1865. It is a United Nations Specialized Agency, currently with over 190 members.

The ITU has three primary functions:

- allocations and use of the radio-frequency spectrum;
- telecommunications standardization;
- development and expansion of worldwide telecommunications.

According to these three functions the ITU have three sectors:

- the Radiocommunications Sector (ITU-R), responsible for frequency allocations and use of the radio-frequency spectrum;
- the Telecommunications Standards Sector (ITU-T), responsible for telecommunications standards; and
- the Telecommunications Development Sector (ITU-D), responsible for the development and expansion of worldwide telecommunications.

The specific frequency bands and other regulatory factors for a particular satellite system have two performances:

- *service(s)* to be provided by the satellite system/network; and
- *location(s)* of the satellite system/network ground terminals.

They both together determine the frequency band, or bands, where the satellite system may operate.

4.2. The electromagnetic frequency spectrum

The wavelength of the free space path signal is the principal parameter that determines the interaction effects of the atmosphere, and the resulting link path degradations.

Communications systems employ the electromagnetic frequency spectrum shown in Table 3 and Figure 3 (according to International Telecommunication Union ITU).

Table 3. Electromagnetic frequency spectrum

Range		Frequency	Length of the wave
Extremely low frequency	ELF	3 ÷ 30 Hz	100000 ÷10000 km
Super low frequency	SLF	30 ÷ 300 Hz	10000 ÷1000 km
Ultra low frequency	ULF	300 ÷ 3000 Hz	1000 ÷100 km
Very Low frequency	VLF	3 ÷ 30 kHz	100 ÷10 km
Low frequency	LF	30 kHz ÷ 300 kHz	10 km ÷1 km
Medium frequency	MF	300 kHz ÷ 3 MHz	1 km ÷100 m
High frequency	HF	3 MHz ÷ 30 MHz	100 m ÷10 m
Very high frequency	VHF	30 MHz ÷ 300 MHz	10 m ÷ 1 m
Ultra high frequency	UHF	300 MHz ÷ 3 GHz	1 m ÷ 10 cm
Super high frequency	SHF	3 GHz ÷ 30 GHz	10 cm ÷ 1 cm
Extremely high frequency	EHF	30 GHz ÷ 300 GHz	1 cm ÷ 1 mm
Terahertz	THz	300GHz ÷3000GHz	1 mm ÷ 0.1 mm
Infrared	IR	3000GHz ÷ 400THz	0,1mm÷0,75μm
Visible		400 THz ÷ 750 THz	0,75μm÷0,4μm
Ultraviolet	UV	750 THz ÷ 3 PHz	0,4μm÷0,1μm

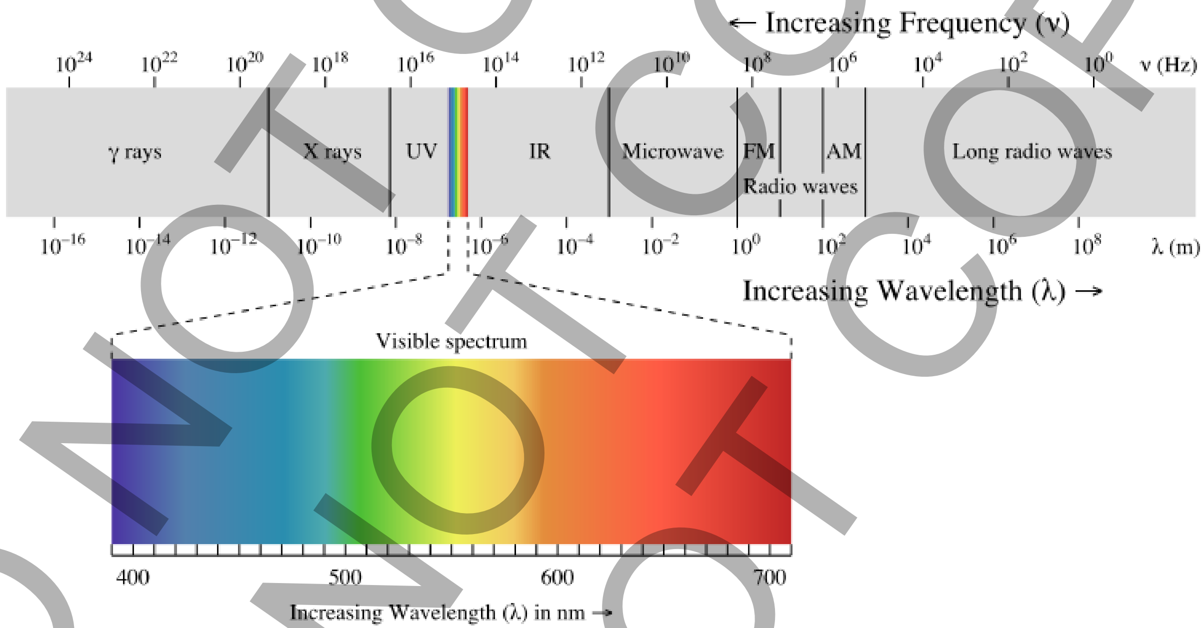


Figure 3. Electromagnetic frequency spectrum ($\lambda = {}^{c}/_{\nu} \approx {}^{3.10^{8}\, m/s}/_{1/s}$)

The frequencies used for satellite communications are accommodated between Super high frequency and Extremely high frequency bands (radio waves) as shown in Table 4.

Table 4. Range of frequencies used for satellite communications

Range	Carrier, GHz	Direction	Service	Bandwidth, MHz
L	1,5-1,6	Downlink	Mobile	100
	1,6-1,7	Uplink	Mobile	100
S	2,5-2,6	Downlink	Broadcast	100
C	3,4-4,2	Downlink	Fixed	800
	4,5-4,8	Downlink	Fixed	300
	5,9-7	Uplink	Fixed	1100
X	7,2-7,7	Downlink	Army	500
	7,9-8,4	Uplink	Army	500

Ku	10,7-11,7	Downlink	Fixed	1000
	11,7-12,5	Downlink	Broadcast	800
	12,5-12,75	Downlink	Fixed (trade)	250
	12,75-13,25	Uplink	Fixed (trade)	250
	14-14,8	Uplink	Fixed	800
	17,3-18,3	Uplink	Fixed	1000
Ka	17,7-20,2	Downlink	Fixed	2500
	20,2-21,2	Downlink	Mobile	1000
	22,5-23	Downlink	Broadcast	500
	27-30	Uplink	Fixed	3000
	30 - 31	Uplink	Mobile	1000

Most commercial communications satellites now operate with a $500MHz$ bandwidth both on the uplink and on the downlink. They use a frequency spectrum $6/4GHz$ and $14/12GHz$ (uplink/downlink). The typical $500MHz$ satellite bandwidth can be segmented into many satellite transponders bandwidths. For example, eight transponders can be provided, each with a $54MHz$ nominal bandwidth and center-to-center frequency spacing $61MHz$. The C-band Fixed Satellite Services (FSS) operate with a 500MHz bandwidth. A typical design would accommodate 12 transponders, each with a bandwidth of 36 MHz, with guard bands of 4MHz between each (Figure 4). A typical commercial communications satellite today can have 24 to 48 transponders, operating in the C-band, Ku-band, or Ka-bands.

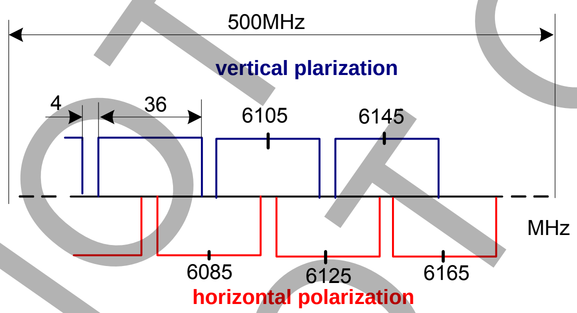


Figure 4. A 500MHz bandwidth accommodate 12 transponders, each with a bandwidth of 36 MHz, with guard bands of 4MHz

Modern communications satellites also employ *polarization frequency reuse* to increase (doubled) the number of transponders: two carriers are used at the same frequency, but with orthogonal polarization. Both linear polarization (horizontal and vertical sense) and circular polarization (right-hand and left-hand sense) have been used.

5. Signal processing elements in satellite communications

Figure 5 shows a the basic signal processing elements in satellite communications and the techniques available for the provision of baseband formatting, source combining, and carrier modulation in satellite communications systems.

Acronyms:

SSB/SC single sideband suppressed carrier

DSB/SC double sideband suppressed carrier

FDM frequency division multiplexing

TDM time division multiplexing
FM/FDM frequency division multiplex
SCPC single channel per carrier
MCPC multiple channel per carrier
FSK frequency shift keying
BPSK binary phase shift keying
QPSK quadrature phase shift keying
QAM quadrature amplitude modulation
FDMA frequency division multiple access
TDMA time division multiple access
CDMA code division multiple access
NIC nearly instantaneous compounding
CVSD continuously variable slope delta modulation
ADPCM adaptive differential pulse code modulation

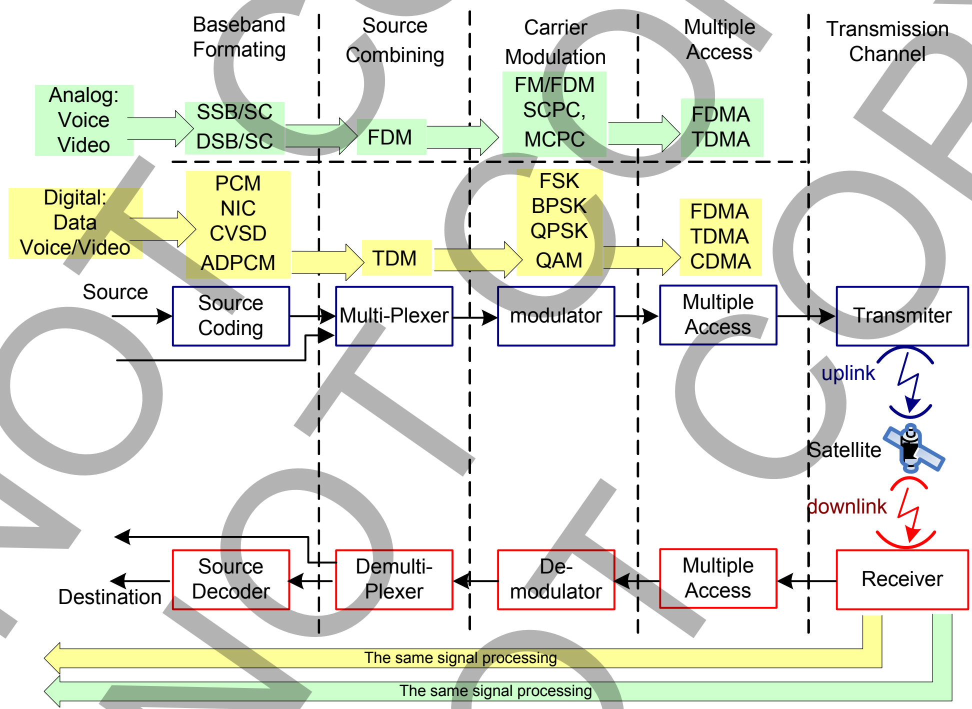


Figure 5. The basic signal processing elements in satellite communications

The Source information may be analog or digital. The first three elements (baseband formatting, source combining, and carrier modulation) prepare the signal for introduction to the transmission channel (the ground-to-satellite-to-ground RF channel). The received signal at the Destination location is subjected to a reverse sequence of processing.

5.1. Analog Systems

In early generation satellites analog transmission has dominated satellite communications since its inception. Even today many satellite systems still transmit telephony and television signals using frequency modulation.

Baseband Formatting - Analog Signals

Analog baseband voice spectrum is typically reduced to the band 300–3400 Hz for electronic transmission. The voice signals are placed on sub-carriers to allow for propagation through the network (figure 6). The sub-carrier format is either Single

Sideband Suppressed Carrier (SSB/SC) or Double Sideband Suppressed Carrier (DSB/SC).

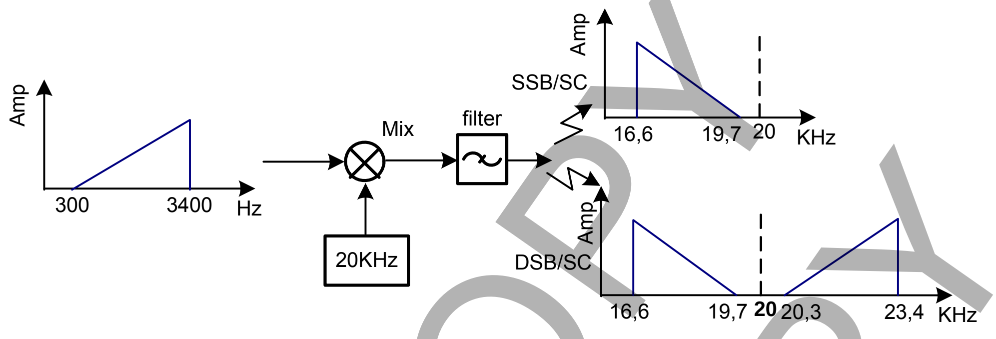


Figure 6. Analog voice baseband formats

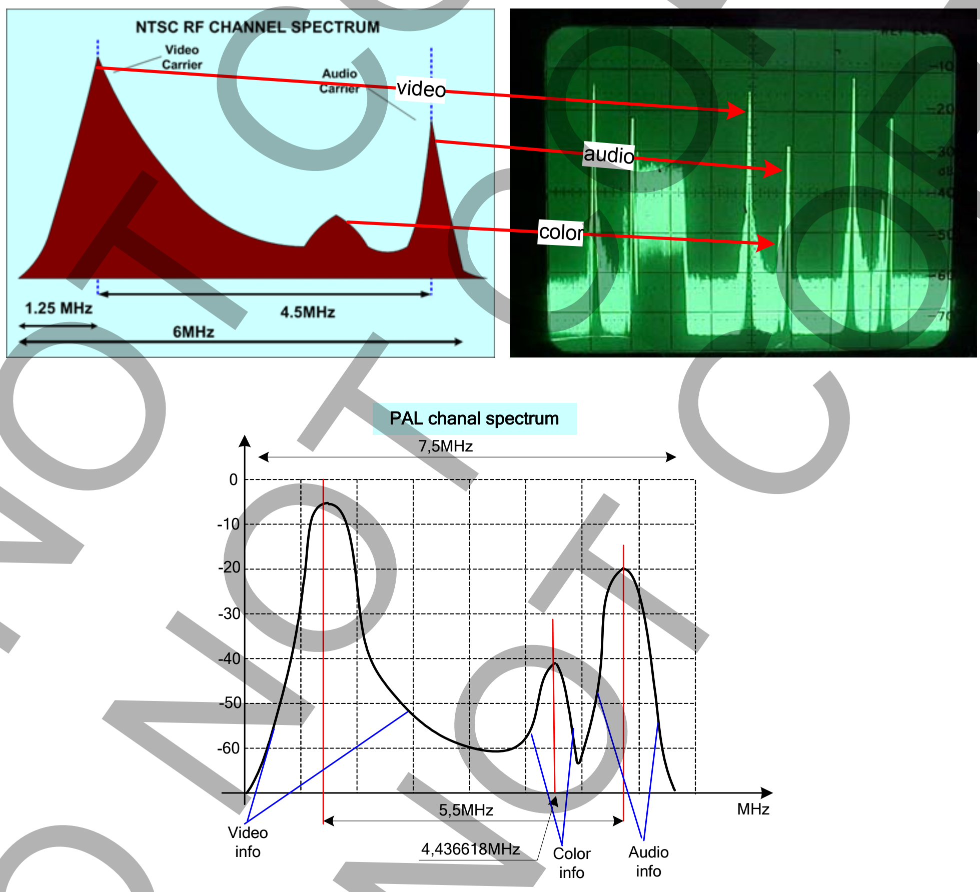


Figure 7. Analog video NTSC and PAL composite baseband signal spectrum

Analog video baseband consists of a composite signal that includes video, color, and audio information (figure 7). The specific format of the components depends on the standard employed for color sub-carrier modulation. Three standards are in use: NTSC (National Television System Comm.), PAL (Phase Alternation Line), SECAM. A modulation format for the satellite channel is a frequency modulation (FM). The composite NTSC signal and all audio channel sub-carriers are combined and the resulting signal frequency modulates the RF carrier, resulting in a total RF spectrum of approximately 36 MHz. It is shown in figure 8.

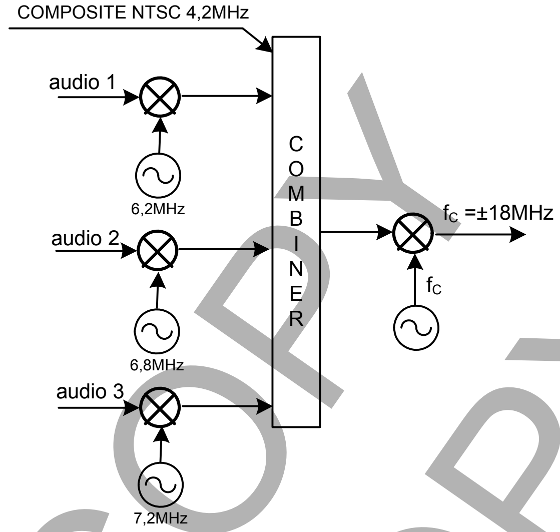


Figure 8. Signal processing format and spectrum for analog video satellite transmission

Source Combining - Analog Signals

Source combining involves the combining of multiple sources into a single signal, which then modulates an RF carrier for transmission through the communications channel. The preferred combining method for analog data is Frequency Division Multiplex (FDM).

The ITU-T FDM standard for FDM is shown in figure 9.

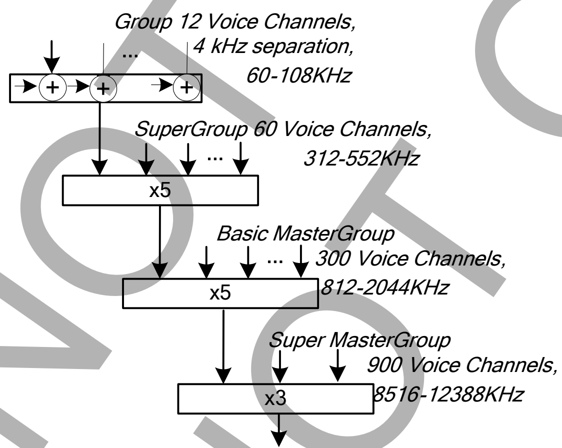


Figure 9. ITU-T FDM standard for analog voice frequency division multiplexing

Carrier Modulation - Analog Signals

Amplitude modulation (AM) was the first method to carry communications on an RF carrier. AM is still used in satellite communications for voice and data communications. Amplitude modulation is produced by mixing the information signal with the RF carrier in a product modulator (mixer) providing a modulated RF carrier where the amplitude envelope is proportional to the information signal.

Suppressed carrier amplitude modulation is the preferred AM modulation implementation for satellite communications. Both *single sideband suppressed carrier AM* (AM-SSB/SC) and *double sideband suppressed carrier AM* (AM-DSB/SC) are used for subcarrier components in satellite communications.

Another technique to modulate carrier of analog signal is Frequency modulation (FM). FM was used extensively in satellite communications for telephony and video transmissions on early generation analog based systems, many of which are still in use.

5.2. Digital Signals

Baseband Formatting

Digital signals dominate satellite communications systems, for data, voice, imaging, and video applications. Digital formatted signals allow for more comprehensive processing capabilities regarding coding, error correction, and data reformatting. The basis for digital communications is the binary digital (2-level) format. There are many of binary waveforms used for the encoding of baseband data - unipolar NRZ and polar NRZ, polar RZ, split phase (Manchester) coding, alternate mark inversion (AMI) and etc.

A measure the quality of digital signals is Bit Error Rate (BER). The bit error rate or bit error ratio is the number of bit errors divided by the total number of transferred bits during a studied time interval. BER is an unitless performance measure, often expressed as a percentage number. The BER is often expressed as a function of the E_b / N_0, (energy per bit to noise power spectral density ratio).

A second step in digital baseband formatting is multi-level coding, where the binary bit stream is combined into groups, called ***symbols***. It reduces the required bandwidth. When two consecutive bits are combined, they are forming a group of two bits, there are four possible combinations of the two bits, resulting in *quaternary encoding*. If three consecutive bits are combined, forming a group of three bits, there are eight possible combinations, resulting in 8-level encoding.

The number of possible levels for “m-ary signal” is

$$m = 2^{N_b} \tag{1}$$

The number of bits per symbol is

$$N_b = \log_2 m \tag{2}$$

The symbol duration is N_b times the bit duration T_b, i.e.,

$$T_S = T_b.N_b \tag{3}$$

The symbol rate is

$$R_S = \frac{R_b}{N_b} \tag{4}$$

where R_b is the bit rate.

The symbol rate is often expressed in units of baud, i.e., the *baud rate*. A transmission rate of $R_S\ simbols/s$ is the same as a rate of $R_S\ baud$. The baud rate is equal to the bit rate only for $N_b = 1$, that is, for binary signals.

The second possibility is when the original source data is analog (voice or video). Conversion to digital form is required before digital formatting is performed. The most popular baseband formatting technique for analog source data is *Pulse Code Modulation* (PCM).

In addition to conventional PCM there are some other digital voice source coding techniques used for the communications systems, as:

Adaptive differential PCM (ADPCM)

ADPCM also uses differential encoding, but takes the mean-square value in the sampling process. It requires fewer coding bits than PCM.

Adaptive delta modulation (ADM) or continuously variable slope delta modulation (CVSD)

ADM uses differential encoding – only changes are transmitted. ADM provides acceptable voice at 24–32 kbps, providing a more spectral efficient option.

Nearly instantaneous compounding (NIC)

NIC achieves bit-rate reduction by taking advantage of short-term redundancy in human speech. It can achieve data rates approaching 1/2 that required for PCM, because of a more efficient use of the frequency spectrum.

Source Combining

Digital satellite communications systems use *Time division multiplex* (TDM). This technique combines multiple digitally encoded signals into a composite signal at a bit rate equal to or greater than the sum of the input rates. Multiple PCM bit streams are combined in a TDM multiplexer, which generates a TDM composite bit sequence that drives the RF modulator (Figure 10). Because PCM samples the incoming signals 8000 times per second, each sample occupies $1/8000s$ ($125\mu s$). According recommendation G.702 ITU-T TDM is organized into a well structured *hierarchy*.

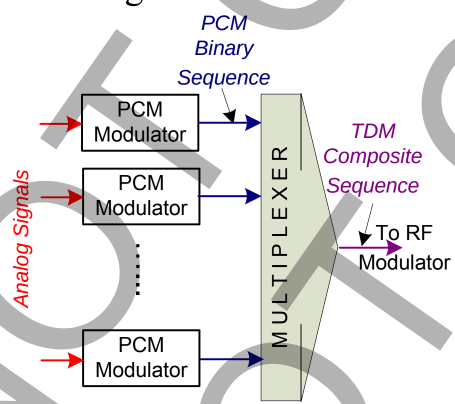


Figure 10. Time division multiplexing source combining process

Two TDM standards for voice circuits are in global use: *DS or T-carrier TDM signalling* and *CEPT TDM signalling*. Each hierarchy starts with 64 kbps analog voice, but the subsequent TDM levels consist of different combinations, as shown in Figure 12. The systems used in Europe and North America are different (Figure 12). The North American standard is based on a 24-channel PCM system, whereas the European system is based on 30/32 channels. This system contains 30 speech channels, a synchronisation channel and a signalling channel, and the gross line bit rate of the system is 2.048 Mbps (32 x 64 Kbps). The system can be adapted for *common channel signalling*, providing 31 data channels and employing a single synchronisation channel.

The following details refer to the European system (Figure 11). The 30/32 channel system uses a *frame* and *multiframe* structure, with each frame consisting of 32 pulse channel time slots numbered 0-31. Slot 0 contains the *Frame Alignment Word* (FAW) and *Frame Service Word* (FSW). Slots 1-15 and 17-31 are used for digitised speech (channels 1-15 and 16-30 respectively). In each digitised speech channel, the first bit is used to signify the polarity of the sample, and the remaining bits represent the amplitude of the sample. The duration of each bit on a PCM system is $0{,}488\mu s$. Each time slot is therefore $3{,}904\mu s$ ($8bits*0{,}488\mu s$). Each frame therefore occupies $125\mu s$ ($32*3{,}904\mu s$).

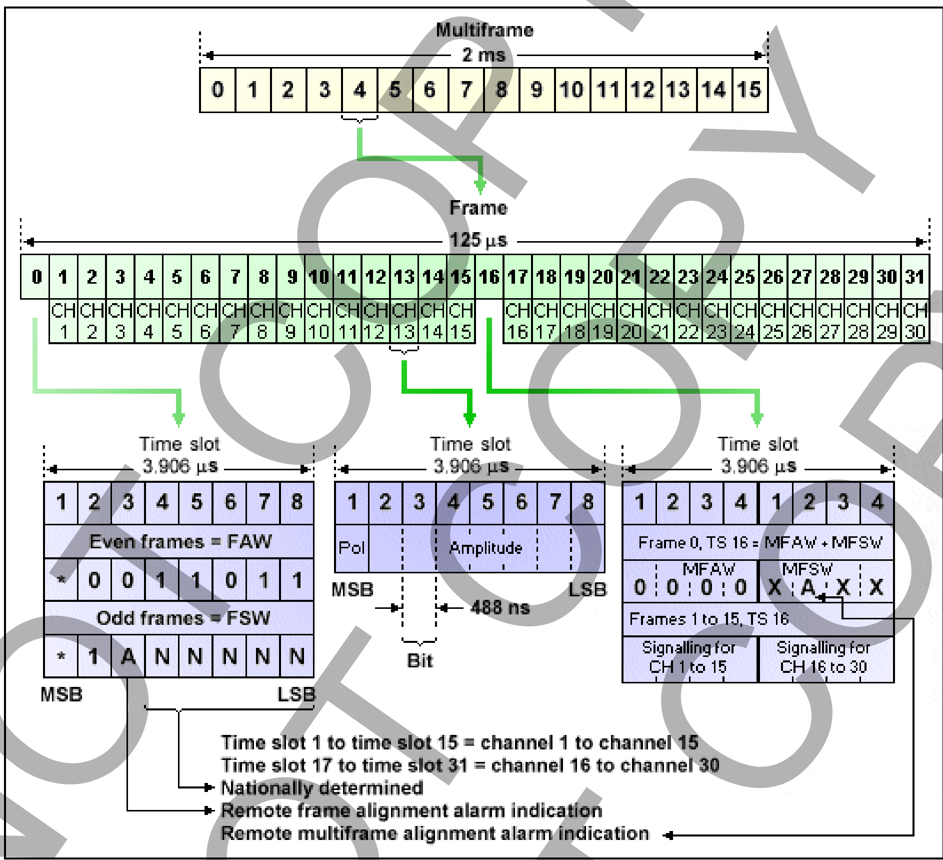


Figure 11. The frame and multiframe structures for a 30/32 channel PCM system

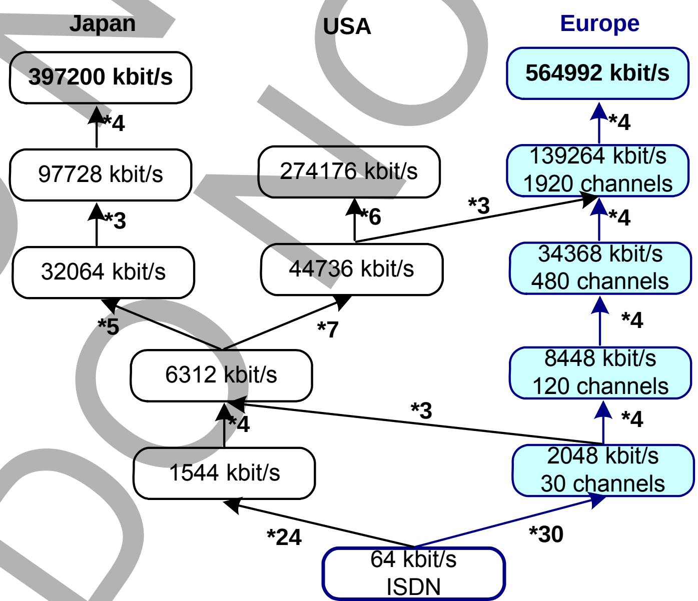


Figure 12. Standardized TDM structures

In order for signalling information (dial pulses) for all 30 channels to be transmitted, the multiframe consists of 16 frames numbered 0-15. In frame 0, slot 16 contains the Multiframe Alignment Word (MFAW) and Multiframe Service Word (MFSW). In frames 1-15, slot 16 contains signalling information for two channels. The frame and multiframe structure are shown below. The duration of each multiframe is $2ms$ $(125\mu s * 16)$. A TDM multiplexer can provide Bit multiplexing, Byte (8 bits symbol) multiplexing and block multiplexing.

Forward Error Correcting (FEC)

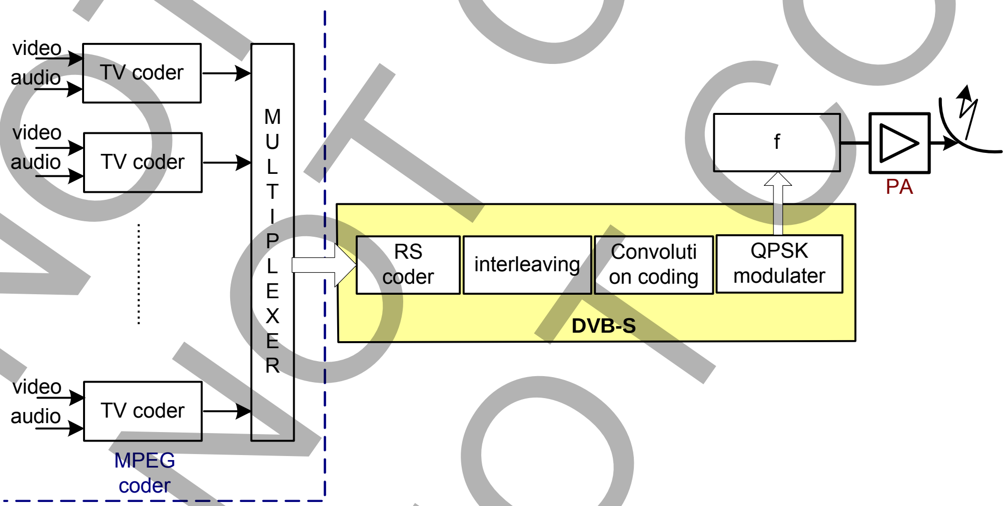
The purpose of Forward Error Correcting (FEC) is to improve the capacity of a transmission channel by adding to the source data redundant information. It is a process known as *channel coding*. Channel coding consists:

Reed Solomon coding;

Interleaving;

Convolution code.

The place of channel coding is between multiplexing and modulating process, as shown in Figure 13.



Фиг. 13. Transmitter digital sat TV.

Channel coding is described in more detail in lecture 3.

Carrier Modulation - Digital Signals

Digital modulation is accomplished by amplitude, frequency, or phase modulation of the carrier by the binary (or m^{-ary}) bit stream.

The major modulation formats used in satellite communications are:

- Binary Phase Shift Keying (BPSK) - polarity changes in binary signal used to produce 180° carrier phase change;

- Differential Phase Shift Keying (DPSK) – phase shift keying where carrier phase is changed only if current bit *differs* from preceding bit. A reference bit must be sent at start of message for synchronization;

- Quadrature Phase Shift Keying (QPSK) – phase shift keying for a 4-symbol waveform. Data bit streams are converted to two bit streams, I and Q, and then binary phase shifted as in BPSK. The adjacent phase shifts are equi-spaced by 90^0. BPSK only requires one-half the bandwidth of BPSK;

- M-ary Phase Shift Keying (MPSK) – phase shift keying for m^{-ary} symbol waveform;

- Minimum Shift Keying (MSK) – phase shift keying with additional processing to smooth data transitions, resulting in reduced bandwidth requirements;

- Quadrature Amplitude Modulation (QAM) – multilevel (higher than binary) modulation; it is a combination of amplitude and phase modulations.

BPSK and QPSK are the most widely used in satellite systems.

In the simplest case PSK is based on phase switching signal to 180^0 when changing from a logic “0” to “1” and “1” to “0”, but the amplitude and frequency remain constant. The change of phase state is finding relatively simple. It is shown in figure 14.

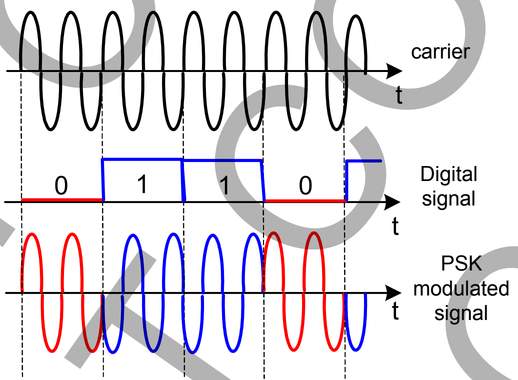


Figure 14. PSK modulation

Quadrature Phase Shift Keying (QPSK)

In QPSK the phase of the carrier is shifted between four positions that are 90 degrees apart. The binary data stream is converted into 2-bit symbols, which are used to phase modulate the carrier. The serial-to-parallel process used to generate the two parallel encoded bit streams is shown in figure 15. Odd numbered bits in the original data sequence $p(t)$ are sent to the ***i*** (in-phase) channel, producing the sequence $p_i(t)$. Even numbered bits are sent to the ***q*** (quadrature) channel, producing the sequence $p_q(t)$. The bit duration is doubled in the ***i*** and ***q*** channels, reducing the bit rate to ½ the original data bit rate (Figure 16).

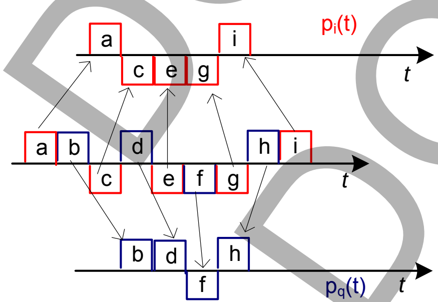


Figure 15. QPSK modulator

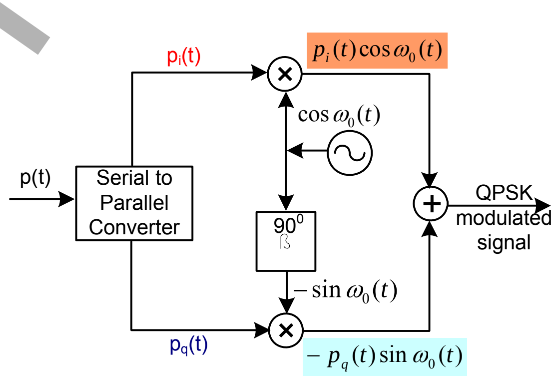


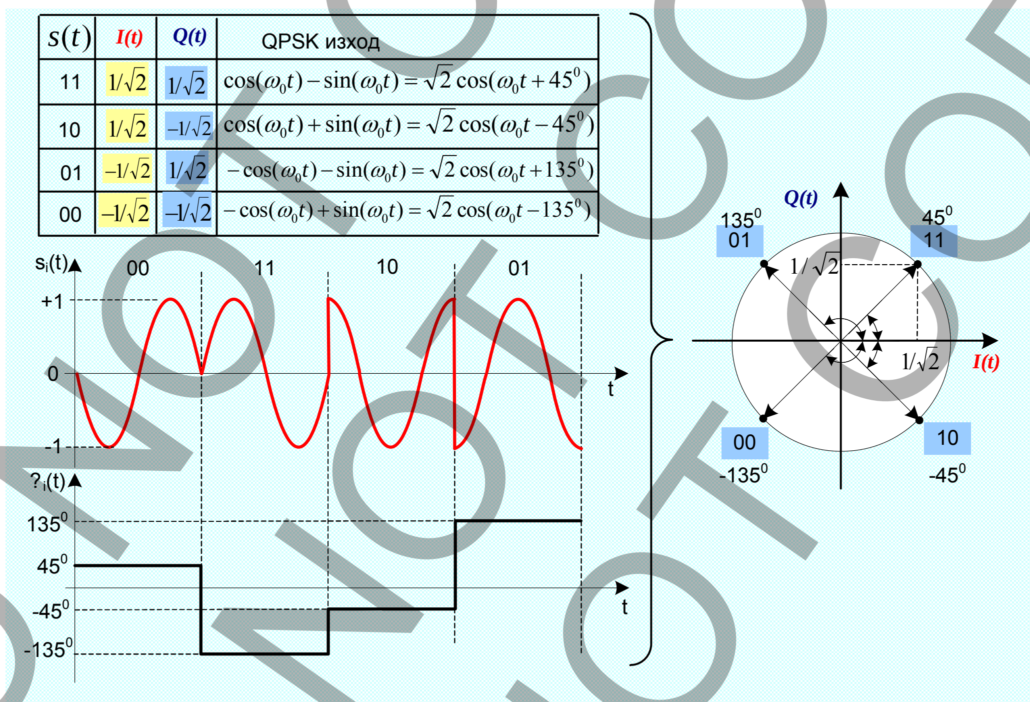
Figure 16. Generation of the QPSK waveform

The in-phase signal is mixed directly with the carrier frequency $\cos(\omega_0 t)$, while the quadrature signal is mixed with a 90^0 phase shifted carrier ($\cos(\omega_0 t + 90^0) = -\sin(\omega_0 t)$). The output of the two mixers is summed to produce the modulator output signal:

$$s(t) = p_i(t)\cos(\omega_0 t) - p_q(t)\sin(\omega_0 t) \qquad (5)$$

The phase state of $s(t)$ depend on the bit values that compose the in-phase and quadrature component signals. Figure 17 and Table 5 show the four possible combinations of bits and the resulting $s(t)$ and the phase state diagram for the four bit sequences, plotted $p_i(t)$ and $p_q(t)$. The symbol phases are orthogonal, 90^0 apart from each other, with one in each quadrant.

The symbol rate is two bit per symbol, $E_s = 2E_b$, and QPSK requires half the transmission bandwidth of BPSK.



$s(t)$	$I(t)$	$Q(t)$	QPSK изход
11	$1/\sqrt{2}$	$1/\sqrt{2}$	$\cos(\omega_0 t) - \sin(\omega_0 t) = \sqrt{2}\cos(\omega_0 t + 45^0)$
10	$1/\sqrt{2}$	$-1/\sqrt{2}$	$\cos(\omega_0 t) + \sin(\omega_0 t) = \sqrt{2}\cos(\omega_0 t - 45^0)$
01	$-1/\sqrt{2}$	$1/\sqrt{2}$	$-\cos(\omega_0 t) - \sin(\omega_0 t) = \sqrt{2}\cos(\omega_0 t + 135^0)$
00	$-1/\sqrt{2}$	$-1/\sqrt{2}$	$-\cos(\omega_0 t) + \sin(\omega_0 t) = \sqrt{2}\cos(\omega_0 t - 135^0)$

Figure 17. The four possible combinations of bits and the resulting $s(t)$ and the phase state diagram

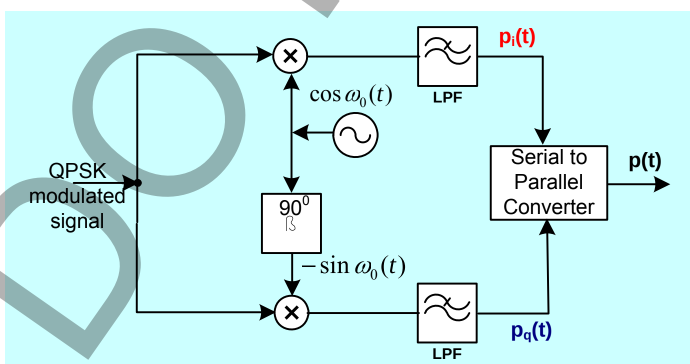


Figure 18. QPSK demodulator

The QPSK demodulator reverses the process as shown in Figure 18. The input signal, which has passed through the communications channel, is split into two channels. Each channel is essentially a BPSK demodulator and the demodulation process produces the in-phase and quadrature components $p_i(t)$ and $p_q(t)$, which are then parallel-to-serial converted to produce the original data stream $p(t)$.

Higher Order Phase Modulation

Higher Order Phase Modulations can achieve further reduction in symbol rate, and the required transmission bandwidth. For example, *8-phase shift keying* (8PSK), which combines groups of three bits per symbol, requires a transmission channel bandwidth of 1/3 BPSK, with the phase state diagram as shown in Figure 19.

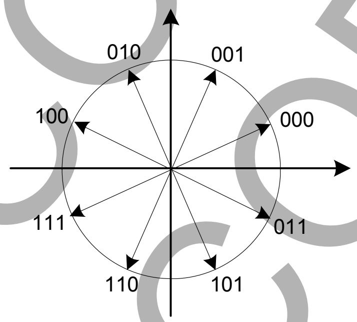


Figure 19. 8PSK phase state diagram

Disadvantage of 8PSK is requirement of twice the power over BPSK or QPSK to achieve the same overall performance. 8PSK Modulation is important in satellite communications systems because the additional bit in the symbol can be used for error correction coding, allowing an additional approximately 3 dB of coding gain. BER values exceeding 1.10^{-10}.

Multiple Access

A multi-user system achieves a specific point in the capacity region depending on how the multi-user channel is shared by the users, which depends on the multiple access technique or Multiple Access (MA). It is the last technique before signal transmitting.

Satellite links are designed to provide desired link availability for average conditions. Satellite MA techniques interconnect ground stations through multiple satellite transponders with the goal of optimizing several system attributes such as: spectral efficiency; power efficiency; reduced latency; and increased throughput.

The MA methods available to the satellite system designer can be categorized into three fundamental techniques:

- Frequency Division Multiple Access (FDMA);
- Time Division Multiple Access (TDMA);
- Code Division Multiple Access (CDMA).

FDMA systems consist of multiple carriers that are separated by *frequency* in the transponder. The transmissions can be analog or digital, or combinations of both. In TDMA the multiple carriers are separated by *time* in the transponder, presenting only one carrier at any time to the transponder. TDMA is most practical for digital data only, because the transmissions are in a burst mode to provide the time division capability. CDMA is a *combination of both frequency and time separation*. It is the most complex technique, requiring several levels of synchronization at both the transmission and reception levels. CDMA is implemented for digital data only, and offers the highest power and spectral efficiency operation of the three fundamental techniques.

Multiple Access is described in more detail in lecture 4.

References

1. A.C. Clarke, "Extraterrestrial Relays",*Wireless World*, Vol. 51, pp. 305 308, October 1945.

2. Bernard Sklar, Digital Communications: Fundamentals and Applications, Second Edition, Prentice-Hall, 2001

3. Dennis Roddy, Satellite communications, McGraw-Hill Professional, 2001 - 569 p.

4. Deffebach H. L. and Frost W. O., A survey of digital baseband signaling techniques. NASA Technical Memorandum NASATM X-64615, June, 1971.

5. Gérard Maral, Michel Bousquet, Satellite communications systems: systems, techniques, and technology, John Wiley and Sons, 2002 – 757p.

6. Ippolito, Louis J., Satellite communications systems engineering: atmospheric effects, satellite link design, and system Performance, 2008 JohnWiley & Sons Ltd, 396 p.

7. International Telecommunications Union, www.itu.int.

8. Korn I., Digital Communications. Van Nostrand Reinhold Company, New York, 1985.

9. Lathi B. P., Modern digital and analog communications systems, Third Edition, Oxford University Press, 1998, 781 стр.

10. Michael O. Kolawole, Satellite communication engineering, Marcel Dekker, 2002 - 263 p.

11. Rappaport T. S., Wireless Communications, Chapter 3 and 4, Prentice Hall, Upper Saddle River, New Jersey, 1996.

12. Zhili Sun, Satellite networking principles and protocols, John Wiley & Sons, 2005. 342 p.

Satellite Communications Systems
Lecture 2.
Link Budget Calculation and Analysis

The goal of satellite communications is to provide links between fixed terminals, mobile terminals and both fixed and mobile terminals on lend, on the air, and at sea.

A link budget is the sum of all gains and losses in the radio connection between two terminals from end to end, including antenna's, feed lines and the path between the antenna's, but also the relevant portions of the transmitter and the receiver.

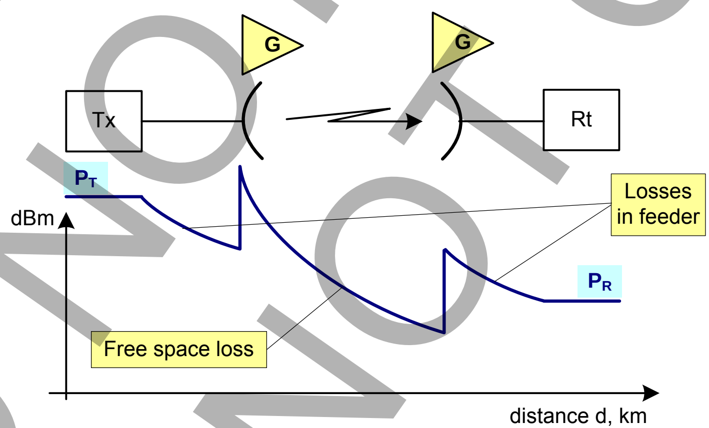


Figure 1. Gains and losses in the radio connection between two terminals

1. The quality of signal transmission

The quality of digital signals (digital links) is measured by the Bit Error Rate (BER). The bit error rate or bit error ratio is the number of bit errors divided by the total number of transferred bits during a studied time interval. BER is a unitless performance measure.

The BER is often expressed as a function of the E_b / N_0, (energy per bit to noise power spectral density ratio).

$$BER = \frac{1}{2} erfc(\sqrt{E_b/N_0}), \quad (1)$$

where the function $erfc$ is called a complimentary error function, it is tabulated.

The relation between BER and E_b/N_0 is shown in Figure 2.

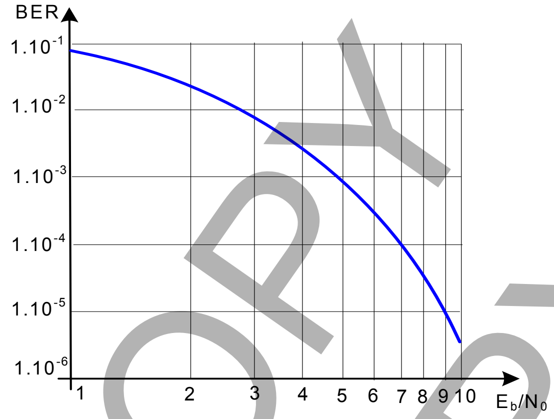


Figure 2. The relation between BER and E_b/N_0

Higher E_b/N_0 means better quality.

The parameter E_b/N_0 is the most common to compare the digital communication systems, even they have differing bit rates or modulation. The quantity E_b is a measure of bit energy, a ratio of the average signal power and bit rate:

$$E_b = \frac{P_{avg}}{R_b}, \quad \frac{W}{bit/s}. \tag{2}$$

Example 1: A signal has power 15W, his bit rate is 200bps. What is the bit energy in decibels?

$$E_b = 10\log 15 - 10\log 200 = -11{,}25dB$$

The quantity N_0 is called the *noise power density* (or noise power spectral density). It is the total noise power in the frequency band of the signal divided by the bandwidth of the signal:

$$N_0 = \frac{P_N}{B_N}, \quad \frac{W}{Hz} \tag{3}$$

Example 2: A signal has a noise power 3W, the signal bandwidth is 500 Hz, $E_b = -11{,}25dB$. What is N_0 and E_b/N_0?

$$N_0 = 10\log 3 - 10\log 500 = -22{,}22dB, \quad \frac{E_b}{N_0} = (-11{,}25) - (-22{,}22) = 10{,}97dB$$

The ratio (3) shows, that if bandwidth of the signal decreases noise power density increases (the same amount of noise occupies a smaller signal space) and the parameter E_b/N_0 decreases. It also decreases when the bit rate increases.

The quality of analog signals is measured by the carrier to noise ratio C/N or by the carrier to noise power spectral density ratio C/N_0. The ratio C/N is the carrier power in the whole useable bandwidth, where C/N_0 is the carrier power per unit bandwidth.

The relation between C/N, C/N_0 and E_b/N_0 is the power of carrier $\boxed{C = E_b.R_b}$.

$$\frac{C}{N_0} = \frac{E_b}{N_0}.R_b, \qquad \text{in dB: } \left(\frac{C}{N_0}\right)_{dB} = \frac{E_b}{N_0} + R_b, \tag{4}$$

$$\frac{C}{N} = \frac{E_b}{N_0}.\frac{R_b}{B}, \qquad \text{in dB: } \left(\frac{C}{N}\right)_{dB} = \frac{E_b}{N_0} + R_b - B, \tag{5}$$

2. Elements of the link

A link consists of three parts with specific parameters: transmitter, receiver and medium, as shown in figure 3.

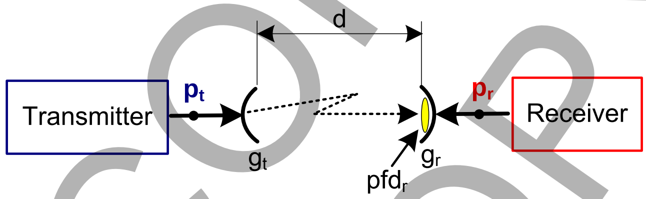


Figure 3. The three parts of link.

The parameters of the link are:

p_T - transmitted power (in watts);

p_R - received power (in watts);

g_T - transmit antenna gain;

g_R - receive antenna gain, and

d - path distance (km).

2.1. Effective Isotropic Radiated Power

An important parameter in the evaluation of the RF link is the ***effective isotropic radiated power*** (EIRP). The EIRP, using the parameters introduced in Figure 3, is defined as

$$EIRP = p_T.g_T, \; W, \tag{6}$$

$$EIRP = 10\log P_T + 10\log G_T, \quad dBW, \tag{7}$$

where p_T - transmitted power (in watts); g_T - transmit antenna gain.

EIRP evaluate the energy potential of the transmitter.

Requirements for the values of EIRP are usually

- $EIRP = 50 \div 100\, dBW$ of the ground transmitter;
- $EIRP = 45 \div 65 dBW$ of the satellite transponder.

Equations (6) and (7) do not account Transmitter Feeder loss L_C and Antenna Pointing Loss L_{Pointing} due to incorrect orientation of the antenna to the satellite.

For precise calculations equation (7) is:

$$EIRP = 10\log P_T + 10\log G_T - 10\log L_C - 10\log L_{\text{Pointing}} \quad dBW, \tag{7a}$$

2.2. Power flux density

A specific parameter is the power flux density (power density p_{PFD}). It is a measure of energy that can be gathering from a particular source (figure 3) and it is expressed in $watts/m^2$.

$$p_{PFD} = \frac{p_T g_T}{4\pi d^2} = \frac{EIRP}{4\pi d^2}, W/m^2, \quad (8)$$

where d - path distance (km).

$$P_{PFD} = EIRP - 10\log\left(4\pi d^2\right) = EIRP - 20\log(d) - 10\log(4\pi) = \\ = EIRP - 20\log(d) - 10{,}99\ dB \quad (9)$$

The power flux density gives the function between transmitted and received power. The power flux density limited service area of the satellite:

- for Direct Broadcast System (DBS), like TV-Sat, TDF, Olimpus $P_{PFD} \geq -103 dBW/m^2$;
- for multifunctional systems (Astra, Eutelsat, Intelsat) $P_{PFD} \geq -(113 \div 116) dBW/m^2$.

Example 3: The transmitted power of Earth station is $10W$, a distance to satellite is $d = 37500km$. The antenna gain is $50dB$. What is the power flux density?

$$EIRP = 10\log P_T + 50 = 10 + 50 = 60, \quad dBW$$

$$P_{PFD} = EIRP - 10\log(37500)^2 - 10{,}99 = 60 - 20\log(37500) - 10{,}99 = \\ = 60 - 91{,}48 - 10{,}99 = -42{,}47dB$$

2.3. Antenna Gain

Diagram of the broadcast antenna means that compared with an isotropic radiator, it concentrates more electromagnetic energy in a definitely direction then to another. In this sense antenna gives "gain" the signal in that direction. The gain of the antenna shows how much more is the gain compared to isotropic antenna gain.

In physical antennas some energy is reflected away by the structure, and some energy is absorbed by lossy components (feeds, struts, subreflectors). In this reason to account for this, an *effective aperture* A_{eff} is defined in terms of an *aperture efficiency* η such that $A_{eff} = \eta . A$, (A is antenna area).

$$g = \eta . \left(\frac{\pi . D}{\lambda}\right)^2 = \frac{4.\pi . A_{eff}}{\lambda^2}, \quad (10)$$

$\eta = 0{,}55 \div 0{,}65$; $\lambda = c/f$ (wave length is the distance from one wave point to the next wave point); c - speed of the light, $c = 3.10^8\ m/s$; f is the frequency.

$$G = 10\log\left[\eta . \left(\frac{\pi . D}{\lambda}\right)^2\right] = 10\log \eta . (10{,}47 . f . D)^2,\ dBi \quad (11)$$

Table 1 gives example for the values of gain depending on frequency and antenna's diameter D.

Table 1. The values of antenna gain

f, GHz	17,5	11,5	3,9	2,6	0,72
D, m \ λ, m	0,017	0,026	0,0,77	0,115	0,417
0,6	**38,7**	**35,0**	**25,5**	**22,1**	**10,9**
2,0	**49,1**	**45,4**	**36,0**	**32,5**	**21,3**
5,0	**57,1**	**43,9**	**43,9**	**30,5**	**29,3**

Figure 4 shows a typical directional antenna pattern for a circular parabolic reflector antenna. The antenna pattern shows the gain as a function of the distance from the boresight direction. The boresight direction is a direction of maximum gain, for which the value g is determined from (11) equation. The 1/2 *power beamwidth* is the contained conical angle α for which the gain has dropped to 1/2 the value at boresight, the power is 3 dB down from the boresight gain value.

The antenna beamwidth for a parabolic reflector antenna can be approximately determined in degree from

$$\alpha \cong \frac{75.\lambda}{D} = \frac{22,5}{D.f}, \text{ in degree}. \qquad (12)$$

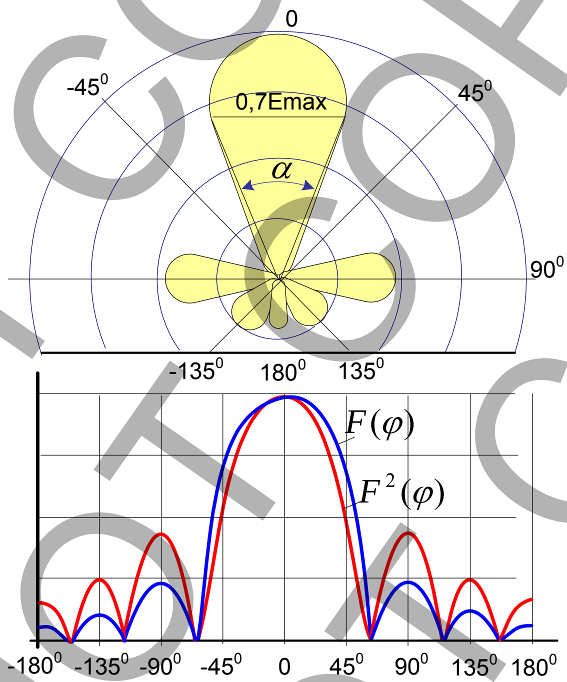


Figure 4. The typical antenna bandwidth

Most antennas have *sidelobes*. Its are regions where the gain may increase due to physical structure elements or the characteristics of the antenna design. It is also possible that some energy may be present behind the physical antenna reflector. Sidelobes are a possible source for noise and interference, for example, when the satellite ground antenna is located near to other antennas or sources of power in the same frequency band as the satellite link.

Because of antenna beamwidth for satellite links is very small (much less than $\alpha < 1^0$), it is requiring careful antenna pointing and control to maintain the link.

2.4 Free-Space Path Loss

When signals are transmitting between the transmitting and receiving antennas signal fade in space. This is due to loss of energy dissipation in the propagation. For GEO orbits satellites the attenuation range is from 195 to 213 dB for 4 / 30 GHz.

According to figure 3 the power p_R got by the receiving antenna will be

$$p_R = p_{PFD}.A_{eff}\ , \qquad (13)$$

where $A_{eff} = \dfrac{G_R.\lambda^2}{4\pi}$.

Replacing A_{eff} and p_{PFD} (8) in equation (13),

$$p_R = p_T.g_T.g_R\left(\frac{\lambda}{4\pi d}\right)^2 = \frac{p_T.g_T.g_R}{l_{FS}}, \tag{14}$$

where $l_{FS} = \dfrac{16\pi^2 d^2}{\lambda^2} = \dfrac{(4\pi.f.d)^2}{c^2}$ is a *free space path loss.*

In dB:

$$L_{FS} = 20\log\left(\frac{4.\pi.d}{\lambda}\right) = 20\log\left(\frac{4\pi.f.d}{c}\right) \tag{15}$$

Example 4: The frequency uplink is 6,175GHz, a distance to satellite is $d = 37500km$. What is the free space path loss?

$$L_{FS} = \frac{(4\pi.f.d)^2}{c^2} = 20\log\frac{4\pi.f.d}{c} = 20\log\frac{4.3,14.6,175.10^9.37,5.10^6}{3.10^8} = 199,73\,dB$$

Equation for Received Power (14) in dB

$$P_R = EIRP + G_R - L_{FS},\ dB \tag{16}$$

gives the basic link equation, sometimes referred to as the *Link Power Budget Equation*, for a satellite communications link, and is the design equation from which satellite design and performance evaluations proceed.

Example 5: What is the power flux density p_{PFD} and power of received signal p_R if both the transmit and receive parabolic antennas have a diameter of 3 m, $D = 3m$, the transmit power is 12 watts, $p_T = 12W$, the antenna efficiency is 55% for both antennas, $\eta = 0,55$. The satellite is in a GSO location, with a range of $d = 37500km$. The frequency of operation is 12 GHz.

First have to estimate antenna gain (equation 11)

$$G_{dB} = 10\log\eta.(10,47.f.D)^2 = 10\log 0,55.109,66.144.9 = 48,93,\ dBi$$

Then EIRP is: $EIRP = 10\log P_T + G_T = 10,79 + 48,93 = 59,72\,dB$

The free space path loss, in dB is (Equation 15)

$$L_{FS} = 10\log\frac{(4\pi.f.d)^2}{c^2} = 20\log\frac{4\pi.f.d}{c} = 215,05\,dB$$

The received power, in db, is then found from the Equation (16):

$$P_R = EIRP + G_R - L = 59,72 + 48,93 - 215,05 = -106,4\,dB$$

The received power in watts:

$$P_R = 10^{\frac{-106.4}{10}} = 2,29.10^{-11}\,W$$

The power flux density, in db, is then found from the Equation (9):

$$P_{PFD} = EIRP - 20\log(d) - 10,99 = 59,72 - 20\log 37,5.10^6 - 10,99 = 102,75\ dB.$$

The calculation shows, that the received power is very, very low.

2.5. System Noise

Noise is a parasitic undesired signal, superimposed on the useful signal, causing adverse effects in the processing of signals. According to the physical character, the noises are: Thermal noise, Shot noise and Flicker noise.

Sources of noise and there frequency domains are different, but in the microwave range (above 300 MHz) matter have two main sources. One of them is related to random changes in the currents of the media in transitions of semiconductor structures (shot noise), and the other - with occasional changes voltages, primarily related to the thermal motion of carriers in the volume of in the wires (thermal noise). Thermal noise is called “white noise” and is a basic in RF band.

The noise introduced by each device in the system is quantified by the introduction of an ***equivalent noise temperature*** t_E. It is defined as the temperature of a passive resistor producing a noise power per unit bandwidth.

Solid state theory gives the following value of noise voltage

$$u_N^2 = 4k.t_E.r.b_N, \quad V^2 \tag{17}$$

where u_N is a noise voltage;

r is a resistance, which occurs on the noise voltage;

k is Boltzmann’s constant: $\begin{cases} k = 1{,}37.10^{-23}\, Ws/\deg, \\ k = -198 dBm/K/Hz, \\ k = -228{,}6 dBW/K/Hz \end{cases}$;

t_E equivalent noise temperature in degree Kelvin;

b_N - noise bandwidth.

The noise power is

$$n_N = \frac{u_N^2}{2r} = k.t_E.b_N, W \tag{18}$$

The noise bandwidth is the RF bandwidth of the information-bearing signal – usually it is the filtered bandwidth of the final detector/demodulator of the link.

A noise power density (noise power spectral density) is

$$n_0 = \frac{n_N}{b_N} = k.t_E\ , \quad W/Hz \tag{19}$$

2.5.1. *Noise figure*

The *noise figure* NF is defined by considering the ratio of the desired signal power to noise power ratio at the input of the device, to the signal power to noise power ratio at the output of the device

$$nf = \frac{(p/n)_{IN}}{(p/n)_{OUT}} \tag{20}$$

In other wise

$$nf = \frac{(p/n)_{IN}}{(p/n)_{OUT}} = \frac{\dfrac{p_{IN}}{k.t_0.b}}{\dfrac{g.p_{IN}}{g.k.(t_0+t_E).b}} = \frac{t_0+t_E}{t_0} = 1+\frac{t_E}{t_0} \tag{21}$$

where t_0 is the input reference temperature, usually set at $290K$, and b is the noise bandwidth.

$$NF = 10\log\left(1+\frac{t_E}{t_0}\right), dB \tag{22}$$

The term in brackets, $\left(1+\frac{t_E}{t_0}\right)$, is sometimes referred to as the ***noise factor***, when expressed as a numerical value.

Corollary:

$$n_{out} = \mathrm{nf}.g.n_{IN} \tag{23}$$

$$t_E = t_0.(\mathrm{nf}-1), \quad K \tag{24}$$

$$t_E = t_0.(10^{\frac{\mathrm{NF}}{10}}-1), \quad dB \tag{25}$$

$$n_N = k.t_0.(\mathrm{nf}-1).b_N, W \tag{26}$$

2.5.2. *Noise Temperature*

Noise temperature concern three types of devices: active devices, passive devices, and the receiver antenna system.

Active devices in the communications system are amplifiers and other components (upconverters, downconverters, mixers, active filters, modulators, demodulators, and etc.). They provide an output power that is greater than the input power (increase the signal level and input noise, g>1).

For active devices noise temperature can be expressed by equations (24) and (25).

Passive devices are waveguides, diplexers, filters, and switches. They reduce the level of the power passing through the device (decrease the signal level, g<1).

A passive device is defined by the *loss factor*:

$$l = \frac{p_{IN}}{p_{OUT}} \tag{27}$$

where p_{IN} and p_{out} are the powers into and out of the device, respectively.

The input to the ideal amplifier noise contribution of the passive device will be

$$t_E = t_0.(l-1) = 290.(l-1), \quad K \tag{28}$$

2.5.3. Receiver Antenna Noise

There are two possible ways to introduce noise into the system at the receiver antenna: from the physical antenna structure (main reflector, subreflector, struts, etc.), it is called *antenna losses*, and from the radio path - *radio noise* or *sky noise*.

The antenna loss is usually included as part of the antenna aperture efficiency, $\eta = 0{,}55 \div 0{,}65$, and does not need to be included in link power budget calculations directly.

Radio noise can be introduced into the transmission path from both natural and human induced sources. This noise power will add to the system noise through an increase in the antenna temperature of the receiver.

The natural components present in radio noise on a satellite link are:

- Galactic noise: $\approx 2{,}4K$ for frequencies above about $1GHz$;

- Atmospheric constituents that absorbs the radiowave and will emit energy in the form of noise, like oxygen, water vapor, clouds, and rain (most severe for frequencies above about 10 GHz).

- Extraterrestrial sources - the moon, sun, and planets.

Human sources of radio noise consist of interference noise in the same information bandwidth induced from:

- communications links, both satellite and terrestrial;
- machinery;
- other electronic devices that may be in the vicinity of the ground terminal.

For the antenna pointing to the sky (ground station antenna) the output noise power from the antenna has two components which are represented by the sky temperature, T_{sky}, and the earth temperature T_{earth}.

Sky temperature, T_{sky} is due to noise originating in the atmosphere. It varies with frequency and the elevation angle φ of the antenna. The sky temperature is higher for $\varphi = 0^0$ (antenna pointing to the horizon) because of the longer path of the radiation through the atmosphere. Elevation angles of less than 10° are usually avoided.

Figure 5 show T_{sky} for different frequency ranges[1].

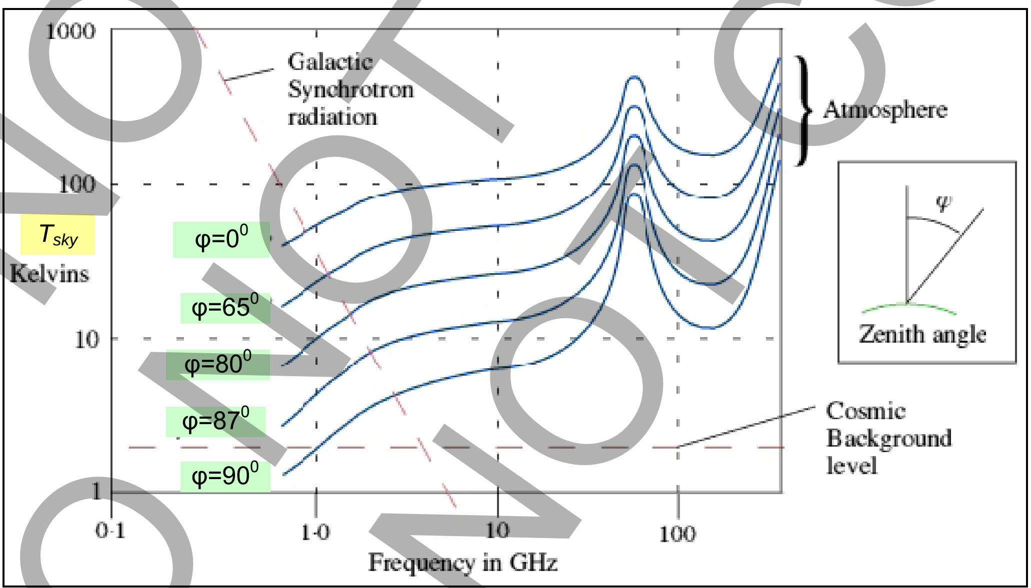


Figure 5. Sky noise for clear air and 7.5 g/m3 of water vapour concentration (φ is the elevation angle)

The earth temperature T_{earth} interference noise enter the system through the sidelobes or backlobes of the ground receiver antenna, it is often difficult to quantify interference noise directly.

For the antenna pointing to the earth (satellite antenna) the noise temperatutre of the antenna is about 290K, the physical temperature of the earth.

[1] J P Silver. Satellite Communications Tutorial, E-mail: john@rfic.co.uk

2.5.4 System Noise Temperature

The noise contributions of each device in the communications transmission path, including sky noise, will combine to produce a total ***system noise temperature***.

Consider a typical satellite receiver system with the components shown in Figure 6: an antenna with a noise temperature of t_A; a low noise amplifier (LNA) with a gain of g_{RF} and noise temperature of t_{RF}; a cable with a line loss of $1/l$ and noise temperature of t_C; a downconverter (mixer) with a gain of g_M and noise temperature of t_M; and finally an intermediate frequency (I.F.) filter and amplifier with a gain of g_{IF} and t_{IF}.

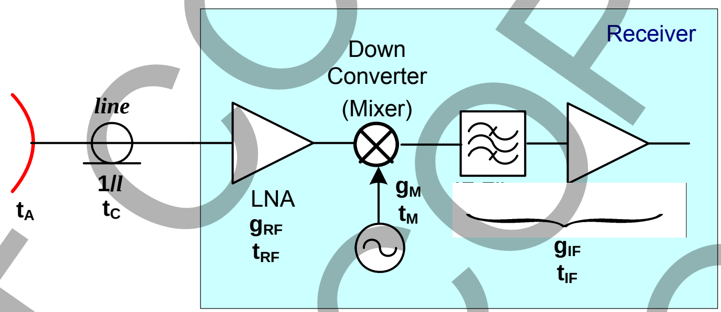


Figure 6. Receiver System Components

The receiver system noise temperature will be

$$t_R = t_{RF} + \frac{t_M}{g_{RF}} + \frac{t_{IF}}{g_{RF} \cdot g_M}, \tag{28}$$

Example 6: A receiver with a low noise amplifier (LNA) with a gain of $G_{RF} = 23dB$ ($g_{RF} = 200$) and noise temperature of $t_{RF} = 50K$; a downconverter (mixer) with a gain of $G_M = -10dB$ ($g_M = 0{,}1$) and noise temperature of $t_M = 500K$, an intermediate frequency (I.F.) unit: $t_{IF} = 1000K$, $G_{IF} = 30dB$ ($g_{IF} = 1000$).What is the receiver system noise temperature.

$$t_R = t_{RF} + \frac{t_M}{g_{RF}} + \frac{t_{IF}}{g_{RF} \cdot g_M} = 50 + \frac{500}{200} + \frac{1000}{200.0{,}1} = 102{,}5K.$$

The total system noise temperature will be

$$t_S = (t_A + t_C)\frac{1}{l} + t_R = \frac{t_A}{l} + t_0\left(1 - \frac{1}{l}\right) + t_R, \tag{29}$$

The system noise temperature t_S represents the noise present at the antenna terminals from all the front-end devices.

2.5.5. The quality of the receiver

The ratio of receiver antenna gain to the receiver system noise temperature G_R / t_S indicates figure of merit.

$$\frac{G}{T} = G_R - T_S,\ dB/K \tag{30}$$

Example 7: Calculate Receiver G/T (dB/K) of a satellite having antenna gain 42 , over all receiver noise temperature $t_S = 75K$.

$$\frac{G}{T} = G_R - 10\log t_S = 42 - 18,75 = 23,25\ dB/K$$

The minimum values for sat TV are:

- $6dB/K$ for an individual receiver;
- $14dB/K$ for a joint receiver.

2.6. Link Performance Parameters

2.6.1. *Carrier-to-Noise Ratio*

The average RF carrier power to noise power ratio C/N is carrier power in the whole useable bandwidth.

The C/N can be expressed in terms of the eirp, G/T, and other link parameters developed earlier

$$\frac{p_R}{n_R} = \left(\frac{C}{N}\right) = \frac{p_T.g_T.g_R}{k.t_S.b_N.(l_{FS}.l_{ad})}; \tag{31}$$

where l_{ad} is sum of all other losses. The other losses could be from the free space path itself, such as rain attenuation, atmospheric attenuation, etc., or from hardware elements such as antenna feeds, line losses, etc.

$$\left(\frac{C}{N}\right) = EIRP + \frac{G_R}{T_S} - (L_{FS} + L_{ad}) - 228,6 - 10\log b_N,\ dB, \tag{32}$$

where the EIRP is in dBW , the bandwidth B_N is in $dBHz$, and $k = -228,6\, dBW/K/Hz$.

The larger the ratio C/N is the better. Typical communications links require minimum C/N values of 6 to 10 dB for acceptable performance. The performance of the link will be degraded in two ways: if the carrier power ***c***, is reduced, and/or if the noise power n_B , increases.

2.6.2. *Carrier-to-Noise Density*

The carrier-to-noise density is defined in terms of noise power density, defined by Equation (19):

$$n_0 = \frac{n_N}{b_N} = \frac{k.t_S.b_N}{b_N} = k.t_S \tag{33}$$

Equation (4) and (5) give the relation between C/N and C/N_0 by the band-width B_N .

$$\frac{c}{n} = \frac{c}{n_0}.\frac{1}{b_N},\quad \left(\frac{C}{N}\right)_{dB} = \left(\frac{C}{N_0}\right) - B_N,\ dB. \tag{34}$$

$$\frac{c}{n_0} = \frac{c}{n}.b_N,\quad \left(\frac{C}{N_0}\right)_{dB} = \left(\frac{C}{N}\right) + B_N,\ dBHz, \tag{35}$$

The carrier-to-noise density behaves similarly to the carrier-to-noise ratio in terms of system performance. The larger value C/N_0 - better performance.

2.6.3. *Energy-Per-Bit to Noise Density*

For digital communications systems, the bit energy, e_b, is more useful than carrier power in describing the performance of the link. The bit energy is related to the carrier power $c = e_b.r_b$ (or $e_b = c.T_b$, T_b is the bit duration).

The ratio e_b/n_0 is related to c/n_0:

$$\left(\frac{e_b}{n_0}\right) = \left(\frac{c}{n_0}\right).\frac{1}{r_b} = \left(\frac{c}{n_0}\right).T_b\,, \tag{36}$$

$$\left(\frac{e_b}{n_0}\right) = \left(\frac{c}{n}\right).\frac{b_N}{r_b}\,, \tag{37}$$

The e_b/n_0 will be numerically equal to the c/n when the bit rate (r_b, bps) is equal to the noise bandwidth (b_N, Hz).

Example 8.[2]:

There is a satellite with a range 40,000 km (range); transmitted power $p_T = 2W$ ($3dB$); antenna gain $g_T = 17\ dB$ (50); channel system noise temperature $t_S = 152K$ and bandwidth $b_N = 500MHz$ of the satellite channel; Frequency = 11GHz.

What are:

a) Power Flux Density p_{PFD} to the received terminal;

b) Received power p_R to the received terminal with antenna effective aperture $A_{eff} = 10m^2$.

c) Received antenna gain g_R;

d) Received ratio c/n

Results:

a) Power Flux Density by (8) and (9):

$$p_{PFD} = \frac{g_T p_T}{4\pi d^2} = \frac{2.50}{4\pi.(4.10^7)^2} = 4{,}97.10^{-15}, W/m^2;$$

$$P_{PFD} = 3 + 17 - 11 - 152 = -143 dBW/m^2$$

b) Received power p_R by (13):

$$p_R = p_{PFD}.A_{eff} = 4.97.10^{-15}.10 = 4.97.10^{-14}\,W$$

$$P_R = 4.97.10^{-14}\,W = -133dBW$$

c) Gain of receiving antenna g_R by (10):

$$\lambda = (c/f) = 3.10^8/11.10^9 = 0.0273,$$

$$G_R = \frac{4\pi.Ae}{\lambda^2} = \frac{4.3{,}14.10}{0{,}0273^2} = \frac{125.6}{7{,}5.10^{-4}} = 167467; \qquad G_R = 52{,}24dB$$

d) Carrier-to-Noise Ratio

It can determinate noise power first

$$p_N = n = k.t_0.b_N = 1{,}37.10^{-23}.152.500.10^6 = 104{,}12.10^{-20},$$

$$P_N = -119{,}8\,dBW.$$

Received power is $P_R = C = 4.97.10^{-14}\,W = -133dBW$;

[2] Mohamed Khedr, Лекция: Satellite Communication Systems, http://webmail.aast.edu/~khedr

Carrier-to-Noise Ratio: $\frac{C}{N} = -133 - (-119.8) = -13.2dBW$.

3. Link Budget

The link budget determines transmitter power needed to obtain sufficient strong signal in the receiver. The quality of the satellite link is estimated by the Carrier-to-Noise Ratio in the receiving earth station or terms of energy per bit to noise density.

Path noise is added to the signal at the uplink and downlink. Path noise is the sum of additive noise effects such as noise caused by atmospheric gases, clouds, rain, depolarization, surface emissions, or extra-terrestrial sources. *Path losses* are introduced in the uplink and the downlink signal paths. Path loss is the sum of signal power losses caused by effects such as gaseous attenuation, rain or cloud attenuation, scintillation loss, angle of arrival loss, or antenna gain degradation.

The total system carrier-to-noise ratio, $(c/n)_S$, is determined by developing the system equations for the total link, including the path degradation parameters.

The parameters used in the link calculations are shown in Figure 7.

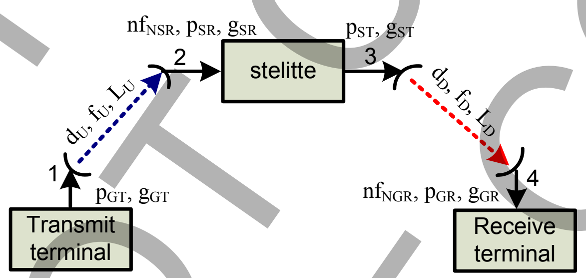


Figure 7. The parameters used in the link calculations

The communications satellite transponder is implemented in one of two general types:

1) the conventional *frequency translation (FT) satellite*, which receives the uplink signal, amplifies, and then reformats and transmits the signal back to the ground received station, and

2) the *on-board processing (OBP) satellite*, which utilizes on-board detection and remodulation to provide two independent communications links (uplink and downlink). Because of that satellite link is assessed in two ways depending on the type of transponder.

3.1. Frequency Translation (FT) Satellite

Figure 8 shows an example of the conventional frequency translation transponder. In the direct transponder, the uplink frequency is converted to the downlink frequency, and after one or more stages of amplification, re-transmitted to the ground. Signal degradations and noise introduced on the uplink are translated to the downlink, and the total performance of the system will be dependent on both links.

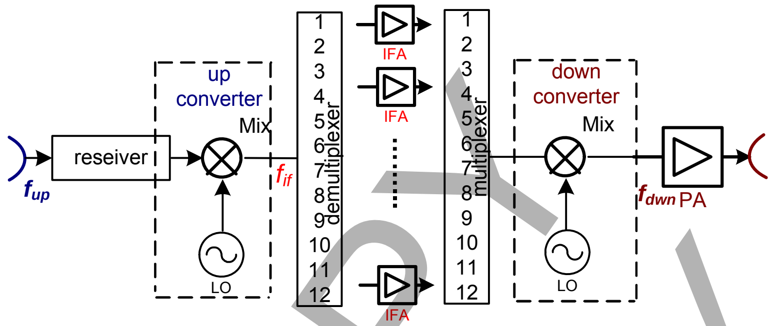


Figure 8. An example of conventional frequency translation transponder

3.1.1. Uplink

To calculate uplink it is starting at the transmit terminal (point 1, Figure 7).

The *eirp* of the ground transmit terminal is:

$$eirp = p_{GT} . g_{GT} \text{ , } W \text{ ,} \tag{38}$$

The carrier power received at the satellite antenna (point 2) is

$$c_{SR} = \frac{p_{GT} . g_{GT} . g_{SR}}{l_{Ufs} . l_{Uad}} \text{ ,} \tag{39}$$

where l_{Ufs} is the uplink free space path loss, a l_{Uad} is the uplink path loss, and g_{GT}, g_{SR} are the transmit and receive antenna gains, respectively.

The noise power at the satellite antenna, point (2), is

$$n_{SR} = k.t_U (1 - \frac{1}{l_{Uad}}).b_U + k.t_{SA}.b_U + k.290.(nf_{NSR} - 1).b_U \text{ ,} \tag{40}$$

where k is Boltzmann's constant, b_U is the uplink information bandwidth, t_{SA} is the satellite receiver antenna temperature, nf_{NSR} is the satellite receiver noise figure, and t_U is the mean temperature of the uplink atmospheric path.

The *uplink* carrier-to-noise ratio, at point (2), is given:

$$\left(\frac{c}{n}\right)_U = \frac{c_{SR}}{n_{SR}} = \frac{p_{GT} . g_{GT} . g_{SR}}{l_{Ufs} . l_{Uad} . k. \left[t_U (1 - \frac{1}{l_{Uad}}) + t_{SA} + 290.(nf_{NSR} - 1) \right] . b_U} \text{ .} \tag{41}$$

This result (41) gives the uplink carrier-to-noise ratio.

3.1.2. Downlink

The procedure is the same that was used for the uplink.

The carrier power received at the received terminal (point 4) is

$$c_{GR} = \frac{p_{ST} . g_{ST} . g_{GR}}{l_{Dfs} . l_{Dad}} \text{ ,} \tag{42}$$

The noise power at the satellite antenna, (point 4), is

$$n_{GR} = k.t_D (1 - \frac{1}{l_{Dad}}).b_D + k.t_{GA}.b_D + k.290.(nf_{NGR} - 1).b_D \text{ .} \tag{43}$$

The *downlink* carrier-to-noise ratio, at point (4), is:

$$\left(\frac{c}{n}\right)_D = \frac{c_{GR}}{n_{GR}} = \frac{p_{ST}.g_{ST}.g_{GR}}{l_{Dfs}.l_{Dad}.k.\left[t_D(1-\frac{1}{l_{Dad}}) + t_{GA} + 290.(nf_{NGR}-1)\right].b_D}. \quad (44)$$

This result (44) gives the downlink carrier-to-noise ratio.

3.1.3. Determination the carrier-to-noise ratio of total link

Consideration should be given to two important conditions specific to the FT satellite:

- the downlink transmits power, p_{ST}, for a frequency translation satellite will contain both the desired carrier component, c_{ST}, and noise introduced by the uplink and by the satellite system itself, n_{ST}:

$$p_{ST} = c_{ST} + n_{ST}; \quad (45)$$

- since there is no on-board processing of the information signal, the satellite input carrier-to-noise ratio must equal the satellite output carrier-to-noise ratio:

$$\left(\frac{c}{n}\right)_{IN} = \left(\frac{c}{n}\right)_{OUT}, \quad (46)$$

that is to say that all noise introduced by the satellite system is accounted for by nf_{SR}.

So, in terms of the link parameters

$$\left(\frac{c}{n}\right)_U = \frac{c_{ST}}{n_{ST}} = \frac{c_{SR}}{n_{SR}} \quad (47)$$

Replacing n_{ST} in Equation (47) with the Equation (45) condition,

$$\left(\frac{c}{n}\right)_U = \frac{c_{ST}}{n_{ST}} = \frac{c_{ST}}{p_{ST}-c_{ST}} = \frac{1}{\frac{p_{ST}}{c_{ST}}-1},$$

consequently

$$\frac{1}{\left(\frac{c}{n}\right)_U} = \frac{p_{ST}}{c_{ST}} - 1. \quad (48)$$

From (48) solving for c_{ST}, and next replacing c_{ST} in Equation (47) with the Equation (45) condition:

$$c_{ST} = \frac{p_{ST}}{1+\frac{1}{\left(\frac{c}{n}\right)_U}}, \qquad \left(\frac{c}{n}\right)_U = \frac{c_{ST}}{n_{ST}} = \frac{p_{ST}-n_{ST}}{n_{ST}} \quad (49)$$

The (49) is transforming and solving for n_{ST},

$$\left(\frac{c}{n}\right)_U = \frac{p_{ST}}{n_{ST}} - 1, \qquad n_{ST} = \frac{p_{ST}}{1+\left(\frac{c}{n}\right)_U}. \quad (50)$$

Next should calculate the necessary carrier power, which must be acceptably the ground station (point 4) for overall link.

The desired carrier power received at the satellite antenna (point 4) is

$$c'_{GR} = \frac{c_{ST} \cdot g_{ST} \cdot g_{GR}}{l_{Dfs} \cdot l_{Dad}}, \tag{51}$$

Replacing c_{ST} from (49) in (51) the result will be

$$c'_{GR} = \frac{p_{ST} \cdot g_{ST} \cdot g_{GR}}{1 + \frac{1}{\left(\frac{c}{n}\right)_U} l_{Dfs} \cdot l_{Dad}}, \tag{52}$$

consequently

$$c'_{GR} = \frac{c_{GR}}{1 + \frac{1}{\left(\frac{c}{n}\right)_U}}, \tag{53}$$

The total noise power received on the ground, n'_{GR}, will be the sum of the noise introduced on the downlink, Equation (43), and the noise transferred from the uplink, Equation (50), and after some transformation

$$n'_{GR} = n_{GR} + \frac{c_{GR}}{1 + \left(\frac{c}{n}\right)_U}. \tag{54}$$

The total carrier-to-noise ratio for the frequency translation transponder is then found as the ratio of

$$\left(\frac{c}{n}\right)_\Sigma = \frac{c'_{GR}}{n'_{GR}}. \tag{55}$$

Replacing n'_{GR} from (54) and c'_{GR} from (53) in (55) can get the total carrier-to-noise ratio

$$\left(\frac{c}{n}\right)_\Sigma = \frac{\dfrac{c_{GR}}{1 + \frac{1}{\left(\frac{c}{n}\right)_U}}}{n_{GR} + \dfrac{c_{GR}}{1 + \left(\frac{c}{n}\right)_U}} \tag{56}$$

After transforming (56) and allude that $\frac{c_{GR}}{n_{GR}} = \left(\frac{c}{n}\right)_D$ result for the FT satellite is:

$$\left(\frac{c}{n}\right)_\Sigma^{FD} = \frac{\left(\frac{c}{n}\right)_U \cdot \left(\frac{c}{n}\right)_D}{1 + \left(\frac{c}{n}\right)_U + \left(\frac{c}{n}\right)_D} \tag{57}$$

If both $\left(\frac{c}{n}\right)_D >> 1$ and $\left(\frac{c}{n}\right)_U >> 1$, equation (57) reduces to

$$\left[\left(\frac{c}{n}\right)_\Sigma^{FD}\right]^{-1} \approx \left[\left(\frac{c}{n}\right)_U\right]^{-1} + \left[\left(\frac{c}{n}\right)_D\right]^{-1} \tag{58}$$

DO NOT COPY
DO NOT COPY

This result is usually acceptable for satellite link analysis, because $\left(\frac{c}{n}\right)_D$ and $\left(\frac{c}{n}\right)_U$ are generally much greater than 1.

3.1.4. Determination the carrier-to-noise density ratio of total link

The total carrier-to-noise density, (c/n_0), can easily be shown to be of the same form as Equation (57)

$$\left(\frac{c}{n_0}\right)_\Sigma^{FD} = \frac{\left(\frac{c}{n_0}\right)_U . \left(\frac{c}{n_0}\right)_D}{1+\left(\frac{c}{n_0}\right)_U + \left(\frac{c}{n_0}\right)_D}, \tag{59}$$

$$\left[\left(\frac{c}{n_0}\right)_\Sigma^{FD}\right]^{-1} \approx \left[\left(\frac{c}{n_0}\right)_U\right]^{-1} + \left[\left(\frac{c}{n_0}\right)_D\right]^{-1}, \tag{60}$$

where

$$\left(\frac{c}{n_0}\right)_U = \frac{p_{GT} . g_{GT} . g_{SR}}{l_{Ufs} . l_{Uad} . k . \left[t_U (1 - \frac{1}{l_{Uad}}) + t_{SA} + 290.(\mathrm{nf}_{NSR} - 1) \right]},$$

$$\left(\frac{c}{n_0}\right)_D = \frac{p_{ST} . g_{ST} . g_{GR}}{l_{Dfs} . l_{Dad} . k . \left[t_D (1 - \frac{1}{l_{Dad}}) + t_{GA} + 290.(\mathrm{nf}_{NGR} - 1) \right]}.$$

3.1.5. Determination the energy-per-bit to noise density

The ratio e_b/n_0 is related to c/n_0:

$$\left(\frac{e_b}{n_0}\right) = \left(\frac{c}{n_0}\right) . \frac{1}{r_b} = \left(\frac{c}{n_0}\right) . T_b, \tag{61}$$

Inserting Equation (61) into Equation (59) the energy-per-bit to noise density can be found:

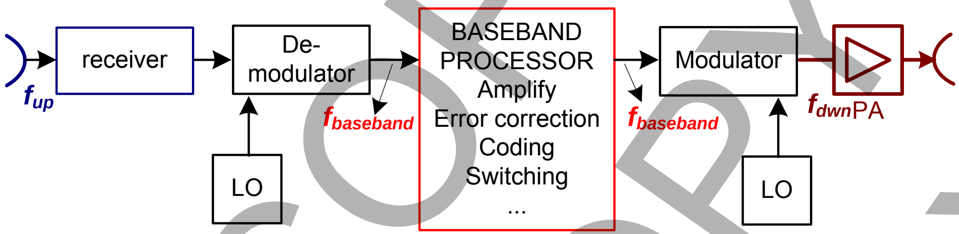
$$\left(\frac{e_b}{n_0}\right)_\Sigma^{FD} = \frac{\left(\frac{e_b}{n_0}\right)_U . \left(\frac{e_b}{n_0}\right)_D}{1+\left(\frac{e_b}{n_0}\right)_U + \left(\frac{e_b}{n_0}\right)_D}, \tag{62}$$

$$\left[\left(\frac{e_b}{n_0}\right)_\Sigma^{FD}\right]^{-1} \approx \left[\left(\frac{e_b}{n_0}\right)_U\right]^{-1} + \left[\left(\frac{e_b}{n_0}\right)_D\right]^{-1}. \tag{63}$$

Equations (62) and (63) give the probability of error for the overall end-to-end digital link.

3.2. The on-board processing (OBP) satellite

The on-board processing satellite utilizes on-board detection and remodulation to provide two independent communications links (uplink and downlink). The information signal on the uplink at a carrier frequency, f_U, after passing through a low noise receiver, is demodulated, and the baseband signal, at f_{BB}, is amplified and enhanced by one or more signal processing techniques. The processed baseband signal is then re-modulated on the downlink, at the carrier frequency, f_D, for transmission to the downlink ground terminal. Figure 9 shows an example of the on-board processing satellite.



Фиг. 9. The on-board processing transponder.

Degradations on the uplink can be compensated for by the on-board processing, and are not transferred to the downlink, consequently, satellites employing on-board processing essentially separate the uplink and downlink. Noise induced on the uplink does not degrade the downlink because the waveform is reduced to baseband and regenerated for downlink transmission.

3.2.1 Uplink and Downlink for OBP satellite

The downlink carrier-to-noise ratio, $(c/n)_D$, or energy-per-bit to noise density, $(e_b/n_0)_D$, for an onboard processing satellite system is essentially independent of the uplink carrier-to-noise ratio. Therefore the link equations (41) and (44) for $(c/n)_U$ and $(c/n)_D$ previously developed for the frequency translation transponder, are applicable to the OBP satellite uplink and downlink.

$$\left(\frac{c}{n}\right)_U^{OBP} = \frac{p_{GT} \cdot g_{GT} \cdot g_{SR}}{l_{Ufs} . l_{Uad} . k . \left[t_U (1 - \frac{1}{l_{Uad}}) + t_{SA} + 290.(\mathrm{nf}_{NSR} - 1) \right] . b_U} . \qquad (64)$$

$$\left(\frac{c}{n}\right)_D^{OBP} = \frac{p_{ST} \cdot g_{ST} \cdot g_{GR}}{l_{Dfs} . l_{Dad} . k . \left[t_D (1 - \frac{1}{l_{Dad}}) + t_{GA} + 290.(nf_{NGR} - 1) \right] . b_D} . \qquad (65)$$

Similarly, it can estimate satellite link by the ratio energy-per-bit to noise density.

$$\left(\frac{e_b}{n_0}\right)_U^{OBP} = \frac{1}{r_U} \frac{p_{GT} \cdot g_{GT} \cdot g_{SR}}{l_{Ufs} . l_{Uad} . k . \left[t_U (1 - \frac{1}{l_{Uad}}) + t_{SA} + 290.(nf_{NSR} - 1) \right]} . \qquad (66)$$

$$\left(\frac{e_b}{n_0}\right)_D^{OBP} = \frac{1}{r_D} \frac{p_{ST} \cdot g_{ST} \cdot g_{GR}}{l_{Dfs} . l_{Dad} . k . \left[t_D (1 - \frac{1}{l_{Dad}}) + t_{GA} + 290.(nf_{NGR} - 1) \right]} . \qquad (67)$$

where r_U and r_D are the uplink and downlink data rates, respectively.

3.2.2 Composite OBP Performance

The overall composite link performance for the OBP satellite is described by its bit error performance, or the *probability of error*, P_E. The overall error performance of the on-board processing transponder will depend on both the uplink and downlink error probabilities. A bit will be correct in the total link if *either* the bit is correct on both the uplink and downlink, *or* if it is in error on both links. The overall probability that a bit is correct, P_{COR}, is therefore[3]

$$P_{COR} = (1-P_U).(1-P_D) + P_U.P_D = 1-(P_U+P_D)+2P_U.P_D, \tag{68}$$

where P_U is the probability of a bit error on the uplink, P_D is the probability of a bit error on the downlink; $(1-P_U)$ is the probability of correct bit on uplink, $(1-P_D)$ is the probability of correct bit on downlink.

The probability of a bit error on the total link is

$$P_E^{OBP} = 1-P_{COR} = P_U + P_D - 2P_U.P_D. \tag{69}$$

The composite link probability of error will be dependent on the uplink and downlink parameters and their impact on the (e_b/n_0) for each link. A specific modulation must be specified to determine the relationship between the bit error probability and the (e_b/n_0) for each link. For example, the procedure for the determination of the composite error performance for an on-board processing transponder system will be demonstrated for a binary frequency-shift keying (BFSK) system with noncoherent detection. The bit error probability is given by Equation (70)[4]

$$P_E = \frac{1}{2}e^{-\frac{1}{2}\left(\frac{E_b}{N_0}\right)} \tag{70}$$

Equations (69) and (70) give the bit error probability for the total link

$$P_E = \frac{1}{2}e^{-\frac{1}{2}\left(\frac{E_b}{N_0}\right)_U} + \frac{1}{2}e^{-\frac{1}{2}\left(\frac{E_b}{N_0}\right)_D} - \frac{1}{2}e^{-\frac{1}{2}\left[\left(\frac{E_b}{N_0}\right)_U + \left(\frac{E_b}{N_0}\right)_D\right]} \tag{71}$$

The ratio e_b/n_0 is related to c/n_0 by (61): $\left(\frac{e_b}{n_0}\right) = \left(\frac{c}{n_0}\right).\frac{1}{r_b} = \left(\frac{c}{n_0}\right).T_b$, and $\left(\frac{c}{n_0}\right)$ can be expressed as $\frac{c}{n_0} = \frac{eirp.g_R}{l_{fs}.l_{ad}.k.t}$, therefore,

$$\left(\frac{e_b}{n_0}\right)_\Sigma = -2.\ln\left\{ e^{-\frac{1}{2.k.r_U}\left(\frac{EIRP_U.\left(g/t\right)_U}{l_{fs}.l_{ad}}\right)_U} + e^{-\frac{1}{2.k.r_D}\left(\frac{EIRP_U.\left(g/t\right)_D}{l_{fs}.l_{ad}}\right)_D} - e^{-\frac{1}{2k}\left[\frac{EIRP_U.\left(g/t\right)_U}{r_U.l_{fs}.L_{ad}} + \frac{EIRP_U.\left(g/t\right)_D}{r_D.l_{fs}.L_{ad}}\right]} \right\} \tag{72}$$

The result (72) is available for the OBP satellite with the binary frequency-shift keying system.

---

[3] Ippolito, Louis J., Satellite communications systems engineering: atmospheric effects, satellite link design, and system Performance, 2008 JohnWiley & Sons Ltd, 396 p.
[4] B. Sklar, Digital Communications – Fundamentals and Applications, Prentice Hall, Englewood Cliffs, NJ, 1988.

References

1. Bernard Sklar, Digital Communications: Fundamentals and Applications, Second Edition, Prentice-Hall, 2001.

2. Broadcom Corporation and Cisco Systems, Digital Transmission: Carrier-to-Noise Ratio, Signal-to-Noise Ratio, and Modulation Error Ratio. http://www.broadcom.com/docs/general/Broadcom-Cisco_CNR-SNR-MER.pdf

3. Dennis Roddy, Satellite communications, McGraw-Hill Professional, 2001 - 569 p.

4. Intuitive Guide to Principles of Communications, Link Budgets, www.complextoreal.com

5. Ippolito, Louis J., Satellite communications systems engineering: atmospheric effects, satellite link design, and system Performance, 2008 JohnWiley & Sons Ltd, 396 p.

6. Mohamed Khedr, Lecture: Satellite Communication Systems, http://webmail.aast.edu/~khedr

7. RPC Telecommunications Ltd., http://www.satcom.co.uk/

8. Silver J. P., Satellite Communications Tutorial, E-mail: john@rfic.co.uk

Satellite Communications Systems

Lecture 3.
Forward Error Correcting in Satellite Communications

1. The necessity of channel coding.

The designation of Source Coding is to get a multiplexing and compressing information. The purpose of Channel Coding is decreasing of errors when signals are transmitted by noisy channel. The capabilities of Channel Coding were investigated by Claude Elwood Shannon (30.04.1916–24.02.2001) and written in "A mathematical theory of communication"[1].

A measure the quality of digital signals is Bit Error Rate (BER). The bit error rate or bit error ratio is the number of bit errors divided by the total number of transferred bits during a studied time interval. BER is a unitless performance measure, often expressed as a percentage number. The BER is often expressed as a function of the E_b / N_0, (energy per bit to noise power spectral density ratio). Figure 1 is shows a comparison of typical coded versus uncoded error performance. If it is needed to provide $BER = 1.10^{-5}$, the uncoded signal needs a ratio $\frac{E_b}{N_0} = 9.4\,dB$, but coded $1/7$ - only $\frac{E_b}{N_0} = 4.6\,dB$. The deference between values of ratio E_b/N_0 coded and uncoded signal is called Coding Gain (G).

$$G = \left(\frac{E_b}{N_0}\right)_U - \left(\frac{E_b}{N_0}\right)_C, dB \qquad (1)$$

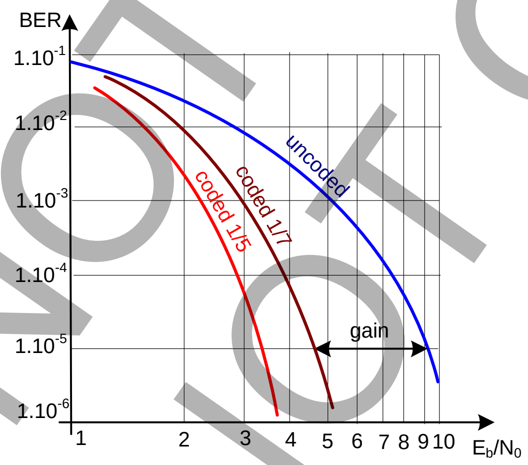


Figure 1. Comparison of typical ceded versus uncoded error performance.

2. Types of error control.

There are two basic methods for error control:

- Error detection and retransmission utilizes parity bits to detect an error (parity bits are redundant bits added to the data);

- Forward Error Correction (FEC) - by parity bits it can detect and correct the errors.

By first method the receiver detect that an error has been made, but does not attempt to correct the error, it requests the transmitter retransmits the information. The first method required two-way link for dialog between the transmitter and receiver.

[1] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech., 1948.

That technique is known as Automatic Repeat Request (ARQ). It can be realize by three ways:

- stop and wait ARQ - the transmitter waits for acknowledgement of each transmission before the next transmission; if some of the message is received with an error, the receiver responds with negative acknowledgement and transmitter repeats this message;
- continuous ARQ with pullback - both the transmitter and receiver work simultaneously, the transmitter is sending messages and receiver is sending acknowledgement data; if some of the message is received with an error, the receiver responds with negative acknowledgement and transmitter retransmits all messages started with the corrupted message to the moment of the negative acknowledgement;
- continuous ARQ with selective repeat - this method is the same as the first, but the transmitter retransmits only the corrupted message.

Which ARQ procedure to choice is a trade-off between the requirements for the efficient of the communication resources.

Therefore, there are developed methods for error correction by encoding the digital signal (ERC error correction coding). ERC has already been used in all types of satellite links. General of the various methods of error correction is the addition of information surplus, called a control feature (additional information) to the basic flow of data through which some errors can be eliminated during signal reception. The receiving system detect and possibly correct errors caused by corruption from the channel. This is called Forward Error Correcting (FEC). The purpose of FEC is to improve the capacity of a transmission channel by adding to the source data redundant information. It is a process known as *channel coding*. Channel coding consists:

Reed Solomon coding;
Interleaving;
Convolution code.

Reed Solomon codes operate on multi-bit symbols rather than individual bits like binary codes. They are a “block” coding technique requiring the addition of redundant parity symbols to the data to enable error correction. The block code operates on a block of bits. Each block is processed as a single unit by both the encoder and decoder. Using a preset algorithm, to a group of bits and it add a coded part to make a larger block. This block is checked at the receiver. The receiver then makes a decision about the validity of the received sequence.

Convolutional codes are referred to as continuous codes as they operate on a certain number of bits continuously.

Interleaving has mitigating properties for fading channels and works well in conjunction with these two types of coding. Standard interleavers scramble code bits among multiple blocks so that they are not contiguous when transmitted; as a result, any bursty errors caused by channel corruption are spread out, into more-random errors after deinterleaving.

For years, convolutional coding with Viterbi decoding has been the predominant FEC technique used in space communications, particularly in geostationary satellite communication networks, such as VSAT (very small aperture terminal) networks. Then, convolutional coding with Viterbi decoding has begun to be supplemented in the satellite communications with Reed-Solomon coding as serially concatenated block and convolutional coding. Typically, the information to be transmitted is first encoded with the Reed-Solomon code, then with the convolutional code. On the receiving end, Viterbi decoding is performed first, followed by Reed-Solomon decoding. This is the technique that is used in most of the direct-broadcast satellite systems and in several of the newer VSAT products.

In 1993, Claude Berrou developed the turbo code, and now it is the most powerful forward error-correction code. Communication systems with a turbo code can approach the theoretical limit of channel capacity, called Shannon Limit. The turbo coder-decoder configuration must include some arrangement of at least two component encoders that are separated by an interleaver. The interleaver in a turbo encoder serves a different purpose than interleavers used by other parts of a communication system. The interleaver in a turbo encoder is designed that the second encoder gets an interleaved version of the same data block that went into the first encoder; thus, the second encoder generates an independent set of code bits.

The place of channel coding is between multiplexing and modulating process, as shown in Figure 2.

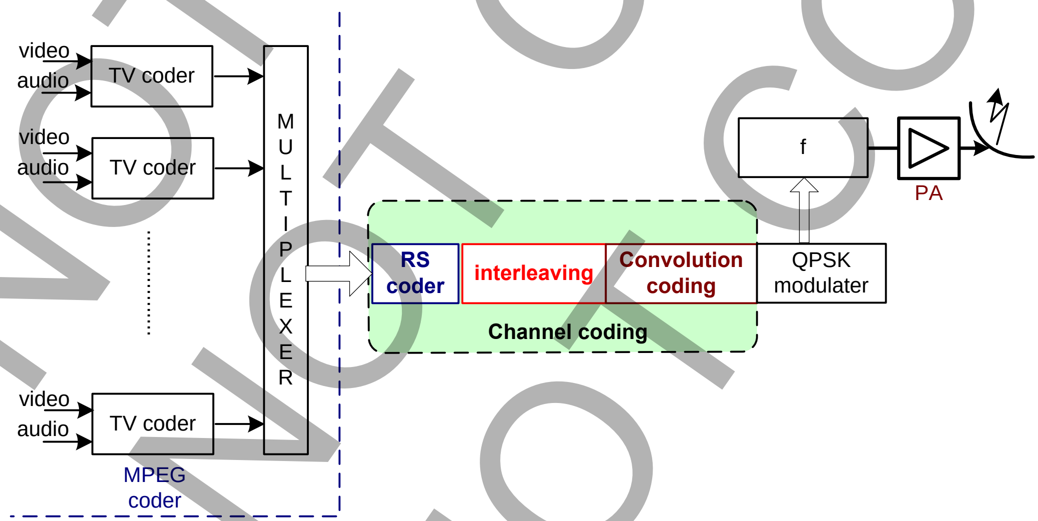


Figure 2. Transmitter digital sat TV.

3. Concept of Hamming Weight and Hamming Distance

The Hamming weight is the largest number of ones in a valid codeword.

Ordinarily, distance is measured by Euclidean concepts such as lengths, angles and vectors. In the digital world, distances are measured between two binary words by the Hamming distance. The Hamming distance is the number of disagreements between two binary sequences of the same size. It is a measure of how apart binary objects are.

Example 1: The Hamming distance between sequences $A = 001$ and $B = 101$ is $d(A,B) = 1$. The Hamming distance between sequences $A = 00101101$ and $B = \bar{1}01\bar{1}\bar{0}101$ is $d(A,B) = 3$. The Hamming weight of a sequence $A = 00101101$ is 4; Hamming weight of a sequence $B = \bar{1}01\bar{1}\bar{0}101$ is 5.

The Hamming distance can be search not only between two binary sequences, but also between all binary words in the same code space. The comparison will give different Hamming distances. The minimum number of symbols that differs the words in the same code space is called the minimum Hamming distance. The number of detected errors depending on the minimum Hamming distance.

$$t = d_{\min} - 1, \quad d_{\min} \geq t + 1, \tag{2}$$

i.e. to detect one error the code has to contain the minimum Hamming distance $d = 2$, to detect two errors - $d = 3$, …

To correct the errors codes have to get more parity bits or symbols. For example, for a sequence of 6 bits, there are 6 different ways to get a 1 bit error and there are 15 different ways we can get 2-bit errors $(\frac{6!}{2!.4!} = 15)$. The number of errors that can be correct is:

$$t = \frac{d_{\min} - 1}{2}, \quad d_{\min} \geq 2t + 1, \tag{3}$$

i.e. to correct one error the code has to contain the minimum Hamming distance $d = 3$.

4. Reed Solomon code

In 1960, Irving Reed and Gus Solomon published a paper in the *Journal of the Society for Industrial and Applied Mathematics*. This paper described a new class of error-correcting codes that are now called *Reed-Solomon* (R-S) *codes*. Reed Solomon codes are linear block nonbinary cyclic codes with symbols made up of *m*-bit sequences, where *m* is any positive integer having a value greater than 2. R-S (*n*, *k*) codes on *m*-bit symbols exist for all *n* and *k* for which

$$k < n < 2^m + 2, \tag{4}$$

where k is the number of data symbols being encoded, and n is the total number of code symbols in the encoded block, $n - k = 2t$ is the number of parity symbols.
For Reed-Solomon codes, the code minimum distance is given by[2]

$$d_{\min} = n - k + 1. \tag{5}$$

If combine Equation (3) and (5) can be express the number of errors that can be correct

$$t = \frac{d_{\min} - 1}{2} = \frac{n - k}{2}, \tag{6}$$

where $\frac{n-k}{2}$ is the largest integer not to exceed $\frac{n-k}{2}$; a Reed–Solomon code can correct up to half as many errors as there are redundant symbols added to the block.

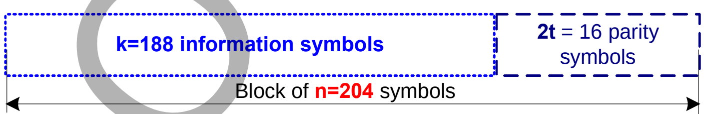


Figure 3. Block of RS (204,188) code - 204 symbols

Reed Solomon codes have information surplus about 8% and $BER = 10^{-4} \div 10^{-11}$.
For example, RS (204,188) code for DVB systems uses 8 bits symbols, and blocks

[2] Gallager, R. G., Information Theory and Reliable Communication, New York: John Wiley and Sons, 1968.

have a length 188 information symbols and 16 parity symbols, as it is shown in Figure 3.

Reed Solomon codes are based on finite fields, called Galois fields (GF) with 2^m elements. Every finite field can be representing by a fixed length binary word. For any number p there exists a finite field $GF(p)$ that contains p elements. It can extend to a field of p^m elements, where *m* is a nonzero positive integer. The new field is called an *extension field* of $GF(p)$, and denoted by $GF(p^m)$. $GF(p^m)$ contains as a subset the elements of $GF(p)$. Symbols from the extension field $GF(p^m)$ are used in the construction of Reed-Solomon (R-S) codes.

The binary field $GF(p)$ is a subfield of the extension field $GF(p^m)$. Besides the numbers 0 and 1, there are additional unique elements in the extension field that will be represented with a new symbol α. Each nonzero element in $GF(p^m)$ can be represented by a power of α. An *infinite* set of elements, *F*, is formed by starting with the elements $\{0,1,\alpha\}$, and generating additional elements by progressively multiplying the last entry by α:

$$F = \{0,1,\alpha,\alpha^2,\alpha^3,..,\alpha^j,...\} = \{0,\alpha^0,\alpha^1,\alpha^2,\alpha^3,..,\alpha^j,...\}. \tag{7}$$

The condition that closes the set of field elements under multiplication is characterized by the irreducible polynomial shown below:

$$\alpha^{(2^m-1)} + 1 = 0$$

or

$$\alpha^{(2^m-1)} = 1 = \alpha^0. \tag{8}$$

Using this polynomial constraint, any field element that has a power equal to or greater than $2^m - 1$ can be reduced to an element with a power less than $2^m - 1$:

$$\alpha^{(2^m+n)} = \alpha^{(2^m-1)}.\alpha^{(n+1)} = \alpha^{(n+1)}. \tag{9}$$

Thus, Equation (8) can be used to form the finite sequence $F*$ from the infinite sequence *F*

$$F* = \{0,1,\alpha,\alpha^2,\alpha^3,..,\alpha^{2^m-2},\alpha^{2^m-1},\alpha^{2^m},...\} = \{0,\alpha^0,\alpha^1,\alpha^2,\alpha^3,..,\alpha^{2^m-2},\alpha^0,\alpha^1,\alpha^2,...\} \tag{10}$$

From Equation (10) the elements of the finite field GF(2*m*) are

$$GF(2^m) = \{0,\alpha^0,\alpha^1,\alpha^2,\alpha^3,..,\alpha^{2^m-2}\}. \tag{11}$$

Each of the 2^m elements of the finite field, $GF(p^m)$, can be represented as a distinct polynomial of degree (m - 1) or less where at least one of the *m* coefficients of $a_i(X)$ is nonzero. The degree of a polynomial is the value of its highest-order exponent.

$$\alpha^i = a_i(X) = a_{i,0} + a_{i,1}X + a_{i,2}X^2 + ... + a_{i,m-1}X^{m-1}, \tag{12}$$

where $i = 0,1,2,...,2^m - 2$.

Example 2: Table 1 shows the possible nine elements of field $GF(2^3)$, $m = 3$, where every row contain values of coefficients $a_{i,0}, a_{i,1}$ *u* $a_{i,2}$ of Equation (12). Seven elements $\{a_i\}$ are differ each other and zero $\alpha^0,\alpha^1,\alpha^2,\alpha^3,\alpha^4,\alpha^5,\alpha^6$, because from (8) $\alpha^{(2^m-1)} = \alpha^{(2^3-1)} = \alpha^0$.

Table 1. Mapping field elements in terms of basis elements for GF(8) with $1+X+X^3$

		Basic elements		
		X^0	X^1	X^2
Elements of the field	0	0	0	0
	α^0	1	0	0
	α^1	0	1	0
	α^2	0	0	1
	α^3	1	1	0
	α^4	0	1	1
	α^5	1	1	1
	α^6	1	0	1
	$\alpha^7=\alpha^0$	1	0	0

One of the benefits of using extension field elements $\{a_i\}$ in place of binary elements is the compact notation that facilitates the mathematical representation of nonbinary encoding and decoding processes. Addition of two elements of the finite field is then defined as the modulo-2 sum of each of the polynomial coefficients of equal powers

$$\alpha^i+\alpha^j=(a_{i,0}+a_{j,0})+(a_{i,1}+a_{j,1})X+\ldots+(a_{i,m-1}+a_{j,m-1})X^{m-1} \qquad (13)$$

Table 2 shows some of primitive polynomials. A class of polynomials called primitive polynomials define the finite fields $GF(p^m)$.

Table 2. Some of primitive polynomials

m		m	
3	$1+X+X^3$	14	$1+X+X^6+X^{10}+X^{14}$
4	$1+X+X^4$	15	$1+X+X^{15}$
5	$1+X^2+X^5$	16	$1+X+X^3+X^{12}+X^{16}$
6	$1+X+X^6$	17	$1+X^3+X^{17}$
7	$1+X^3+X^7$	18	$1+X^7+X^{18}$
8	$1+X^2+X^3+X^4+X^8$	19	$1+X+X^2+X^5+X^{19}$
9	$1+X^4+X^9$	20	$1+X^3+X^{20}$
10	$1+X^3+X^{10}$	21	$1+X^2+X^{21}$
11	$1+X^2+X^{11}$	22	$1+X+X^{22}$
12	$1+X+X^4+X^6+X^{12}$	23	$1+X^5+X^{23}$
13	$1+X+X^3+X^4+X^{13}$	24	$1+X+X^2+X^7+X^{24}$

First of them $f(X)=1+X+X^3$ defines finite field $GF(2^3)$. There are $2^3=8$ elements in the field defined by $f(X)$. Now the roots of $f(X)=0$ must be found. The familiar binary elements, 1 and 0, do not satisfy (are not roots of) the polynomial $f(X)=1+X+X^3$, since $f(1)=1$ and $f(0)=0$ (using modulo-2). A fundamental theorem of algebra states that a polynomial of degree m must have precisely m roots, consequently $f(X)=0$ has three roots, and if α is a root of the polynomial $f(X)$ it is possible to write the following:

$$1+\alpha+\alpha^3=0, \quad \alpha^3=-1-\alpha=1+\alpha, \tag{14}$$

because of $-1=+1$ in the binary field.

Thus, α^3 is expressed as a weighted sum of $\alpha-terms$ having lower orders. In fact all powers of α can be so expressed.

$$\alpha^4=\alpha.\alpha^3=\alpha.(1+\alpha)=\alpha+\alpha^2,$$

$$\alpha^5=\alpha.\alpha^4=\alpha^2+\alpha^3=1+\alpha+\alpha^2$$

$$\alpha^6=\alpha.\alpha^5=\alpha(1+\alpha+\alpha^2)=\alpha+\alpha^2+\alpha^3=1+\alpha^2$$

$$\alpha^7=\alpha.\alpha^6=\alpha(1+\alpha^2)=\alpha+\alpha^3=1=\alpha^0. \tag{15}$$

Therefore the eight finite field elements of $GF(2^3)$ are

$$\{0,\alpha^0,\alpha^1,\alpha^2,\alpha^3,\alpha^4,\alpha^5,\alpha^6\}. \tag{16}$$

4.1. Reed Solomon coding

The generating polynomial for an R-S code in terms of the parameters *n*, *k*, *t*, and any positive integer $m>2$ is

$$g(X)=g_0+g_1X+g_2X^2+...+a_{2t-1}X^{2t-1}+X^{2t}. \tag{17}$$

where $n-k=2t$ is the number of parity symbols, and *t* is the symbol-error correcting capability of the code.

Since the generator polynomial is of degree 2*t*, there must be precisely 2*t* successive powers of α that are roots of the polynomial. We designate the roots of $g(X)$ as $\alpha^1,\alpha^2,\alpha^3,...,\alpha^{2t}$. It is not necessary to start with the root α; starting with any power of α is possible.

For example, there is the (7, 3) R-S code, $t=2$. The generator polynomial in terms of its 2*t* = *n* - *k* = 4 roots can be described, as follows:

$$g(X)=(X-\alpha).(X-\alpha^2).(X-\alpha^3).(X-\alpha^4)=X^4-\alpha^3X^3+\alpha^0X^2-\alpha^1X+\alpha^3$$

$$g(X)=\alpha^3+\alpha^1X+\alpha^0X^2+\alpha^3X^3+X^4. \tag{18}$$

R-S codes are cyclic codes and encoding in systematic form is analogous to the binary encoding procedure. The information message (a polynomial, $m(X)$) can be multiplied by X^{n-k} so that it is right-shifted $n-k$ positions and then appending a parity polynomial, $p(X)$, by placing it in the leftmost $n-k$ stages. Then it can divide $X^{n-k}.m(X)$ by the generator polynomial $g(X)$, which is written in the following form:

$$m(X).x^{n-k}=q(X).g(X)+p(X), \tag{19}$$

where $q(X)$ and $p(X)$ are quotient and remainder polynomials, respectively.

As in the binary case, the remainder $p(X)$ is the parity. Equation (19) can also be expressed as follows:

$$p(X)=m(X).x^{n-k}\oplus g(X). \tag{20}$$

The resulting codeword polynomial, $U(X)$ can be written as

$$U(X)=p(X)+m(X).X^{n-k}. \tag{21}$$

Example 3: There is a message that consist three symbols 010 110 111, and it has to be coded by RS(7,3) code with the generator polynomial (18): $g(X)=\alpha^3+\alpha^1X+\alpha^0X^2+\alpha^3X^3+X^4$.

First, according to table 1 these three symbols are $\alpha^1,\alpha^3,\alpha^5$. The information message polynomial is $m(X)=\alpha^1+\alpha^3X+\alpha^5X^2$, it has to multiply by $X^k=X^4$. The result is $X^4.m(X)=\alpha^1X^4+\alpha^3X^5+\alpha^5X^6$. Next, it has to divide this upshifted message polynomial by the generator polynomial (18) to get a parity polynomial, $p(X)=\alpha^0+\alpha^2X+\alpha^4X^2+\alpha^6X^3$. Then, from Equation (21), the codeword polynomial is:

$$U(X)=\alpha^0+\alpha^2X+\alpha^4X^2+\alpha^6X^3+\alpha^1X^4+\alpha^3X^5+\alpha^5X^6.$$

The circuit to encode a symbol sequence in systematic form with the (n, k) R-S code with the generator polynomial requires the implementation of a linear feedback shift register (LFSR) circuit. The linear feedback shift register is a dynamic memory of binary 0 and 1. The memory consist a number of cages that determines the memory. When in the most left stage enters a new symbol all next shift to the right, and this one that is the rightmost go out of the register.

Figure 4 shows a circuit of LFSR for RS (7,3) code from example 3.

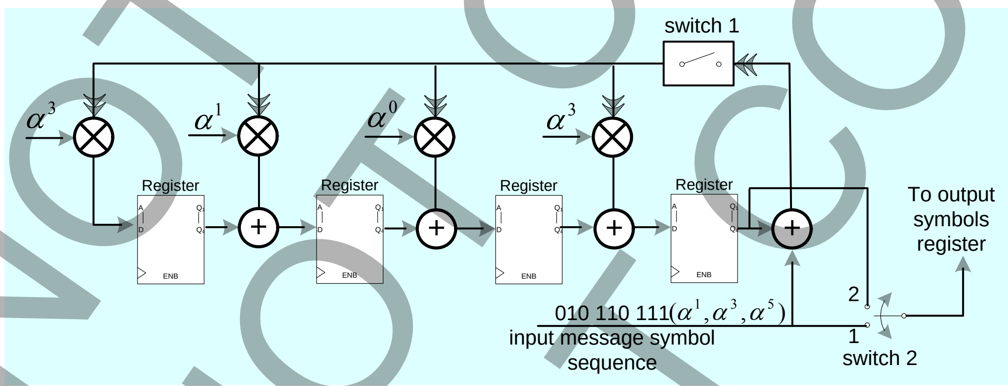


Figure 4. A linear feedback shift register of RS (7,3) code

The encoder forms codewords in the steps, as follows:

1) Switch 1 is closed during the first k clock cycles to allow shifting the message symbols into the $(n-k)-stage$ shift register.

2) Switch 2 is in the position 1 during the first k clock cycles in order to allow simultaneous transfer of the message symbols directly to an output register (not shown in Figure 4).

3) After 3rd clock cycle all message symbols are transferred to the output register, switch 1 is opened and switch 2 is moved to the position 2.

4) The remaining $(n-k)$ clock cycles the parity symbols contained in the shift register are moved to the output register.

5) The total number of clock cycles is equal to n, and the contents of the output register is the codeword polynomial $U(X)=p(X)+m(X).X^{n-k}$, where $p(X)$ is the parity polynomial and $m(X)$ the information message in polynomial form.

During the transfer time the rightmost symbol of the information sequence is the earliest symbol, and the rightmost bit is the earliest bit. Table 3 shows the operational steps during the first $k=3$ shifts of the encoding circuit of Figure 4.

Table 3. The operational steps to shift symbols in the register

входна последователност			такт	съдържание на регистъра				обратна връзка
α^1	α^3	α^5	**0**	0	0	0	0	α^5
	α^1	α^3	**1**	α^1	α^6	α^5	α^1	$\alpha^0\ (\alpha^1 \oplus \alpha^3)$
		α^1	**2**	α^3	0	α^2	α^2	$\alpha^4\ (\alpha^2 \oplus \alpha^1)$
		-	**3**	α^0	α^2	α^4	α^6	-

Explanations:

$\alpha^5.\alpha^3 = \alpha^8 = \alpha^1.\alpha^7 = \alpha^1.1 = \alpha^1$ - first clock, first register;

$\alpha^5.\alpha^1 = \alpha^6$ - first clock, second register;

$\alpha^5.\alpha^0 = \alpha^5.1 = \alpha^5$ - first clock, third register;

$\alpha^5.\alpha^3 = \alpha^8 = \alpha^1.\alpha^7 = \alpha^1.1 = \alpha^1$ - first clock, forth register;

$\alpha^0.\alpha^3 \oplus \alpha^5 = 1.\alpha^3 \oplus \alpha^5 = \alpha^2$ (001) - second clock, first register;

$\alpha^0.\alpha^0 \oplus \alpha^6 = 1.1 \oplus \alpha^6 = 1 \oplus 1 \oplus \alpha^2 = \alpha^2$ (001) - second clock, second register;

$\alpha^0.\alpha^1 \oplus \alpha^1 = 1.\alpha^1 \oplus \alpha^1 = 0$ (000) - second clock, third register;

$\alpha^0.\alpha^3 = 1.\alpha^3 = \alpha^3$ (110) - second clock, forth register;

....

Thus, after the third clock cycle, the register contents are the four parity symbols, $\alpha^0, \alpha^2, \alpha^4, \alpha^6$. Then, switch 1 of the circuit is opened, switch 2 is moved to the position 2, and the parity symbols from the register are shifted to the output. The output codeword, $U(X)$, is:

$$\begin{aligned} U(X) &= \alpha^0 + \alpha^2 X + \alpha^4 X^2 + \alpha^6 X^3 + \alpha^1 X^4 + \alpha^3 X^5 + \alpha^5 X^6 = \\ &= (100) + (001)X + (011)X^2 + (101)X^3 + (010)X^4 + (110)X^5 + (111)X^6 \end{aligned} \qquad (22)$$

4.2. Reed Solomon decoding[3]

The presence of noise in the communication channel causes errors in the information and the received message can be different from original message. For instant, assume that during transmission the codeword (22) becomes corrupted so that two symbols are received in error. Remember that the number of errors corresponds to the maximum error-correcting capability of the code. The receiver will get a message:

$$r(X) = U(X) + e(X), \qquad (23)$$

where $e(X)$ is a polynomial of error.

For this example of seven-symbol codeword, the error pattern, $e(X)$, can be described in polynomial form as:

[3] Bernard Sklar, Digital Communications: Fundamentals and Applications, Second Edition, Prentice-Hall, 2001.

$$e(X) = \sum_{n=0}^{6} e_n . X^n \ . \tag{24}$$

Assume that corrupted symbols are forth and fifth, polynomial of error is:

$$\begin{aligned} e(X) &= 0 + 0X + 0X^2 + \alpha^2 X^3 + \alpha^5 X^4 + 0X^5 + 0X^6 \\ &= 000 + 000X + 000X^2 + 001X^3 + 111X^4 + 000X^5 + 000X^6 \end{aligned} \tag{25}$$

According to Equation (23) the received message will be (using modulo-2 on Equations (22) and (25)):

$$\begin{aligned} r(X) &= 100 + 001X + 011X^2 + 100X^3 + 101X^4 + 110X^5 + 111X^6 = \\ &= \alpha^0 + \alpha^2 X + \alpha^4 X^2 + \alpha^0 X^3 + \alpha^6 X^4 + \alpha^3 X^5 + \alpha^5 X^6 \end{aligned} , \tag{26}$$

because of $101 \oplus 001 = 100 = \alpha^0$ (forth symbol) and $010 \oplus 111 = 101 = \alpha^6$ (fifth symbol).

In this example, there are four unknowns—two error locations and two error values. Therefore, four equations are required for their solution.

The decoding order is as follows:

- Syndrome Computation;
- Error Location;
- Error Values Computation;
- Correcting the Received Polynomial with Estimates of the Error Polynomial.

1) Syndrome Computation

The syndrome is the result of a verification on $r(X)$ to determine whether $r(X)$ is a valid member of the codeword set. The syndrome S has value 0 when $r(X)$ is a member. Any nonzero value of S indicates the presence of errors. The syndrome S is made up of $n-k$ symbols, $\{S_i\}\,(i = 1,...,n-k)$. Thus, for (7,3)RS code, there are four symbols in every syndrome vector; their values can be computed from the received polynomial, $r(X)$. It can be made by the structure of the code

$$U(X) = m(X).g(X) \ . \tag{27}$$

This structure shows that every valid codeword polynomial $U(X)$ is a multiple of the generator polynomial $g(X)$. Consequently, the roots of $g(X)$ must also be the roots of $U(X)$. Thus, each of the roots of $g(X)$ should yield zero $r(X)$ (received message) only when it is a valid codeword. Any errors will give a nonzero result in one or more of the computations. The computation of a syndrome symbol can be described as follows:

$$S_i = r(X)\big|_{X=\alpha^i} = r(\alpha^i), \qquad i = 1,2,...,n-k \ \ (i = 1,2,3,4 \ \textit{for example}) \tag{28}$$

In the example, $r(X)$ contain two corrupted by error symbols (Equation 25). The computation of the four syndrome symbols is:

$$\begin{aligned} S_1 &= r(\alpha) = \alpha^0 + \alpha^3 + \alpha^6 + \alpha^3 + \alpha^{10} + \alpha^8 + \alpha^{11} = \\ &= \alpha^0 + \alpha^3 + \alpha^6 + \alpha^3 + \alpha^2 + \alpha^1 + \alpha^4 = \alpha^3 \end{aligned} \tag{29}$$

$$\begin{aligned} S_2 &= r(\alpha^2) = \alpha^0 + \alpha^4 + \alpha^8 + \alpha^6 + \alpha^{14} + \alpha^{13} + \alpha^{17} = \\ &= \alpha^0 + \alpha^4 + \alpha^1 + \alpha^6 + \alpha^0 + \alpha^6 + \alpha^3 = \alpha^5 \end{aligned} \tag{30}$$

$$S_3 = r(\alpha^3) = \alpha^0 + \alpha^5 + \alpha^{10} + \alpha^9 + \alpha^{18} + \alpha^{18} + \alpha^{23} = \\ = \alpha^0 + \alpha^5 + \alpha^3 + \alpha^2 + \alpha^4 + \alpha^4 + \alpha^2 = \alpha^6 \quad (31)$$

$$S_4 = r(\alpha^4) = \alpha^0 + \alpha^6 + \alpha^{12} + \alpha^{12} + \alpha^{22} + \alpha^{23} + \alpha^{29} = \\ = \alpha^0 + \alpha^6 + \alpha^5 + \alpha^5 + \alpha^1 + \alpha^2 + \alpha^2 = 0 \quad (32)$$

The result shows, that received codeword contains errors.

2) Error Location

Suppose there are v errors in the codeword at location $X^{J_1}, X^{J_2}, ..., X^{J_v}$. Then, the error polynomial $e(X)$ shown in Equations (24) and (25) can be written as follows:

$$e(X) = e_{j_1} X^{J_1} + e_{j_2} X^{J_2} + ... + e_{j_v} X^{J_v}. \quad (33)$$

The indices $1,2,...,v$ refer to the first, second, …, v^{th} errors, and the index j refers to the error location. To correct the corrupted codeword, each error value e_{j_l} and its location X_{j_l}, ($l = 1,2,...,v$) must be determined. An error locator number can be mark as $\beta_l = \alpha^{j_l}$. Next, by substituting α^i into the received polynomial for $i = 1,2,...,2t$ can be obtaining the $n - k = 2t$ syndrome symbols:

$$S_1 = r(\alpha) = e_{j_1}\beta_1 + e_{j_2}\beta_2 + ... + e_{j_v}\beta_v$$
$$S_2 = r(\alpha^2) = e_{j_1}\beta_1^2 + e_{j_2}\beta_2^2 + ... + e_{j_v}\beta_v^2$$
$$...$$
$$S_{2t} = r(\alpha^{2t}) = e_{j_1}\beta_1^{2t} + e_{j_2}\beta_2^{2t} + ... + e_{j_v}\beta_v^{2t} \quad (34)$$

There are $2t$ unknowns (t error values and t locations), and $2t$ simultaneous equations. However, these $2t$ simultaneous equations cannot be solved in the usual way because they are nonlinear (as some of the unknowns have exponents). Any technique that solves this system of equations is known as a *Reed-Solomon decoding algorithm*.

If a nonzero syndrome vector (one or more of its symbols are nonzero) has been computed, that signifies that an error has been received. Next, it is necessary to learn the location of the error or errors. An error-locator polynomial, $\sigma(X)$, can be defined as

$$\sigma(X) = (1 + \beta_1 X)(1 + \beta_2 X)...(1 + \beta_v X) = 1 + \sigma_1 X + \sigma_2 X^2 + ... + \sigma_v X^v. \quad (35)$$

The roots of $\sigma(X)$ are $\frac{1}{\beta_1}, \frac{1}{\beta_2}, ..., \frac{1}{\beta_v}$. The reciprocal values of the roots of $\sigma(X)$ show the error-location numbers of the error polynomial $e(X)$. Then, using autoregressive modelling techniques[4] [7], it can form a matrix from the syndromes, where the first t syndromes are used to predict the next syndrome.

[4] Blahut, R. E., *Theory and Practice of Error Control Codes* (Reading, MA: Addison-Wesley, 1983)

$$\begin{bmatrix} S_1 & S_2 & S_3 \ldots S_{t-1} & S_t \\ S_2 & S_3 & S_4 \ldots S_t & S_{t+1} \\ \ldots & & & \\ S_{t-1} & S_t & S_{t+1} \ldots S_{2t-3} & S_{2t-2} \\ S_t & S_{t+1} & S_{t+2} \ldots S_{2t-2} & S_{2t-1} \end{bmatrix} \begin{bmatrix} \sigma_t \\ \sigma_{t-1} \\ \ldots \\ \sigma_2 \\ \sigma_1 \end{bmatrix} = \begin{bmatrix} -S_{t+1} \\ -S_{t+2} \\ \ldots \\ -S_{2t-1} \\ -S_{2t} \end{bmatrix}. \tag{36}$$

For the (7, 3) RS code, the matrix size is 2 × 2, and the model is written as:

$$\begin{bmatrix} S_1 & S_2 \\ S_2 & S_3 \end{bmatrix} \begin{bmatrix} \sigma_2 \\ \sigma_1 \end{bmatrix} = \begin{bmatrix} S_3 \\ S_4 \end{bmatrix} \tag{37}$$

$$\begin{bmatrix} \alpha^3 & \alpha^5 \\ \alpha^5 & \alpha^6 \end{bmatrix} \begin{bmatrix} \sigma_2 \\ \sigma_1 \end{bmatrix} = \begin{bmatrix} \alpha^6 \\ 0 \end{bmatrix} \tag{38}$$

To solve for the coefficients σ_1 and σ_2 and of the error-locator polynomial, $\sigma(X)$ (35), first can take the inverse of the matrix in Equation (38). The inverse rule of a matrix $[A]$ is:

$$Inv[A] = \frac{cofactor[A]}{\det[A]}.$$

Consequently,

$$\det \begin{bmatrix} \alpha^3 & \alpha^5 \\ \alpha^5 & \alpha^6 \end{bmatrix} = \alpha^3\alpha^6 - \alpha^5\alpha^5 = \alpha^9 + \alpha^{10} = \alpha^2 + \alpha^3 = \alpha^5$$

$$cofactor \begin{bmatrix} \alpha^3 & \alpha^5 \\ \alpha^5 & \alpha^6 \end{bmatrix} = \begin{bmatrix} \alpha^6 & \alpha^5 \\ \alpha^5 & \alpha^3 \end{bmatrix}$$

$$Inv \begin{bmatrix} \alpha^3 & \alpha^5 \\ \alpha^5 & \alpha^6 \end{bmatrix} = \frac{\begin{bmatrix} \alpha^6 & \alpha^5 \\ \alpha^5 & \alpha^3 \end{bmatrix}}{\alpha^5} = \alpha^{-5} \begin{bmatrix} \alpha^6 & \alpha^5 \\ \alpha^5 & \alpha^3 \end{bmatrix} = \alpha^2 \begin{bmatrix} \alpha^6 & \alpha^5 \\ \alpha^5 & \alpha^3 \end{bmatrix} = \begin{bmatrix} \alpha^8 & \alpha^7 \\ \alpha^7 & \alpha^5 \end{bmatrix} = \begin{bmatrix} \alpha^1 & \alpha^0 \\ \alpha^0 & \alpha^5 \end{bmatrix} \tag{39}$$

Verification: If the inversion was performed correctly, the multiplication of the original matrix by the inverted matrix should yield an identity matrix.

$$\begin{bmatrix} \alpha^3 & \alpha^5 \\ \alpha^5 & \alpha^6 \end{bmatrix} \begin{bmatrix} \alpha^1 & \alpha^0 \\ \alpha^0 & \alpha^5 \end{bmatrix} = \begin{bmatrix} \alpha^4 + \alpha^5 & \alpha^3 + \alpha^{10} \\ \alpha^6 + \alpha^6 & \alpha^5 + \alpha^{11} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{40}$$

The coefficients σ_1 and σ_2 of the error-locator polynomial, $\sigma(X)$, are solving from Equation (38)

$$\begin{bmatrix} \sigma_2 \\ \sigma_1 \end{bmatrix} = \begin{bmatrix} \alpha^1 & \alpha^0 \\ \alpha^0 & \alpha^5 \end{bmatrix} \begin{bmatrix} \alpha^6 \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha^7 \\ \alpha^6 \end{bmatrix} = \begin{bmatrix} \alpha^0 \\ \alpha^6 \end{bmatrix} \tag{41}$$

From Equations (41) and (35) it can represent $\sigma(X)$:

$$\sigma(X) = 1 + \sigma_1 X + \sigma_2 X^2 = 1 + \alpha^6 X + \alpha^0 X^2 \tag{42}$$

The roots of $\sigma(X)$ are the reciprocals of the error locations. Once these roots are located, the error locations will be known. Because of each of the field elements can be root of $\sigma(X)$ (if it yields $\sigma(X) = 0$), it has to solve the $\sigma(X)$ polynomial with each of the field elements:

$$\sigma(\alpha^0)=\alpha^0+\alpha^6+\alpha^0=\alpha^6\neq 0$$
$$\sigma(\alpha^1)=\alpha^0+\alpha^7+\alpha^2=\alpha^2\neq 0$$
$$\sigma(\alpha^2)=\alpha^0+\alpha^8+\alpha^4=\alpha^6\neq 0$$
$$\sigma(\alpha^3)=\alpha^0+\alpha^9+\alpha^6=0\Rightarrow error$$
$$\sigma(\alpha^4)=\alpha^0+\alpha^{10}+\alpha^8=0\Rightarrow error$$
$$\sigma(\alpha^5)=\alpha^0+\alpha^{11}+\alpha^{10}=\alpha^2\neq 0$$
$$\sigma(\alpha^6)=\alpha^0+\alpha^{12}+\alpha^{12}=\alpha^0\neq 0$$

According to Equation (35), the error locations are at the inverse of the roots of the polynomial. Therefore $\sigma(\alpha^3)=0$ indicates that one root exits at $1/\beta_l=\alpha^3$. Thus, $\beta_l=1/\alpha^3=\alpha^4$. Similarly, $\sigma(\alpha^4)=0$ indicates that another root exits at $1/\beta_{l'}=\alpha^4$. Thus, $\beta_{l'}=1/\alpha^4=\alpha^3$, where $l,l',...,l^v$ refer to the first, second, …, v^{th} error. Therefore, in this example, there are two-symbol errors, so that the error polynomial is of the following form:

$$e(X)=e_{j_1}X^{J_1}+e_{j_2}X^{J_2} \tag{43}$$

Thus, the two errors were found at locations α^3 and α^4, i.e. $\beta_1=\alpha^{j_1}=\alpha^3$ and $\beta_2=\alpha^{j_2}=\alpha^4$.

2) Error Values Computation

To calculate the error value in the location $\beta_1=\alpha^3$ and $\beta_2=\alpha^4$ can be used any of the four syndrome equations (34). For instance, assume that are S_1 and S_2

$$S_1=r(\alpha)=e_1\beta_1+e_2\beta_2$$
$$S_2=r(\alpha^2)=e_1\beta_1^2+e_2\beta_2^2, \tag{44}$$

where indexes of the errors are 1 and 2, because they are associated with locations $\beta_1=\alpha^3$ and $\beta_2=\alpha^4$.

These equations in matrix form are:

$$\begin{bmatrix}\beta_1 & \beta_2\\ \beta_1^2 & \beta_2^2\end{bmatrix}\begin{bmatrix}e_1\\ e_2\end{bmatrix}=\begin{bmatrix}S_1\\ S_2\end{bmatrix} \tag{45}$$

$$\begin{bmatrix}\alpha^3 & \alpha^4\\ \alpha^6 & \alpha^8\end{bmatrix}\begin{bmatrix}e_1\\ e_2\end{bmatrix}=\begin{bmatrix}\alpha^3\\ \alpha^5\end{bmatrix} \tag{46}$$

To solve for the error values e_1 and e_2, the matrix in Equation (46) is inverted

$$inv\begin{bmatrix}\alpha^3 & \alpha^4\\ \alpha^6 & \alpha^8\end{bmatrix}=\frac{\begin{bmatrix}\alpha^1 & \alpha^4\\ \alpha^6 & \alpha^3\end{bmatrix}}{\alpha^3\alpha^1-\alpha^6\alpha^4}=\frac{\begin{bmatrix}\alpha^1 & \alpha^4\\ \alpha^6 & \alpha^3\end{bmatrix}}{\alpha^4+\alpha^3}=\alpha^{-6}\begin{bmatrix}\alpha^1 & \alpha^4\\ \alpha^6 & \alpha^3\end{bmatrix}=\alpha^1\begin{bmatrix}\alpha^1 & \alpha^4\\ \alpha^6 & \alpha^3\end{bmatrix}= \tag{47}$$
$$=\begin{bmatrix}\alpha^2 & \alpha^5\\ \alpha^7 & \alpha^4\end{bmatrix}=\begin{bmatrix}\alpha^2 & \alpha^5\\ \alpha^0 & \alpha^4\end{bmatrix}$$

It can calculate the error value from Equations (47) and (46)

$$\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} \alpha^2 & \alpha^5 \\ \alpha^0 & \alpha^4 \end{bmatrix} \begin{bmatrix} \alpha^3 \\ \alpha^5 \end{bmatrix} = \begin{bmatrix} \alpha^5 + \alpha^{10} \\ \alpha^3 + \alpha^9 \end{bmatrix} = \begin{bmatrix} \alpha^5 + \alpha^3 \\ \alpha^3 + \alpha^2 \end{bmatrix} = \begin{bmatrix} \alpha^2 \\ \alpha^5 \end{bmatrix} \tag{48}$$

4) Correcting the Received Polynomial with Estimates of the Error Polynomial

Using Equations (43) and (48) the estimated error polynomial is

$$\widehat{e}(X) = e_1 X^{J_1} + e_2 X^{J_2} = \alpha^2 X^3 + \alpha^5 X^4 \tag{49}$$

The algorithm that can repair the received polynomial is

$$\widehat{U}(X) = r(X) + \widehat{e}(X) = U(X) + e(X) + \widehat{e}(X) \tag{50}$$

$$r(X) = 100 + 001X + 011X^2 + 100X^3 + 101X^4 + 110X^5 + 111X^6$$

$$\widehat{e}(X) = 000 + 000X + 000X^2 + 001X^3 + 111X^4 + 000X^5 + 000X^6$$

$$\begin{aligned} \widehat{U}(X) &= 100 + 001X + 011X^2 + \underline{101}X^3 + \underline{010}X^4 + 110X^5 + 111X^6 = \\ &= \alpha^0 + \alpha^2 X + \alpha^4 X^2 + \alpha^6 X^3 + \alpha^1 X^4 + \alpha^3 X^5 + \alpha^5 X^6 \end{aligned} \tag{51}$$

Thus, received information message that is located in the rightmost $k = 3$ symbols is the same as transmitted $010\ 110\ 111 \to \alpha^1\ \alpha^3\ \alpha^5$.

Advantages of RS codes:

- The RS code achieves the largest possible code minimum distance for any linear code with the same encoder input and output block lengths.
- The RS code can be configured with long block lengths (in bits) with less decoding time than other codes of similar lengths; This is because the decoder logic works with symbol-based rather than bit-based arithmetic.
- R-S Codes Perform Well Against Burst Noise.

5. Interleaving

Interleaving is a powerful technique that can be used in digital communications systems to enhance the random error correcting capabilities of block codes such as Reed-Solomon codes to the point that they can be effective in a burst noise environment. Interleaving is helpful when burst errors are present.

The transmitted message can be received with errors because of system's noise environment. Errors which can be classified into three broad categories[5]:

1) *Random errors* - the bit errors are independent of each other. Additive noise typically causes random errors.

2) *Burst errors* - the bit errors occur sequentially in time and as groups.

3) *Impulse errors* - large blocks of the data are full of errors.

Random errors occur in a channel when individual bits in the transmitted message are corrupted by noise. They can be described as some single erroneous bits in the message. Thermal noise in communications channels is the main cause for random errors. Reed-Solomon codes are designed specifically for correcting random errors. If the number of errors per data block is small, (up to 4% byte errors) they can be totally corrected by Reed-Solomon codes.

[5] Interleaving for Burst Error Correction, AHA products groop, Moscow, Idaho, USA, www.aha.com.

Burst errors are characterized as a series of adjacent erroneous bits in a signal. They can be caused by fading in a communication channel and its can be difficult to correct for some codes. However, block codes (in particular Reed-Solomon codes) can handle burst noises effectively. The ability of a block code to correct burst errors depends upon the number of errors in the signal. Interleaving can be used to extend the error correcting capability of the Reed-Solomon code.

Impulse errors can cause catastrophic failures in the communications system. In general, all coding systems fail to reconstruct the message in the presence of catastrophic errors. However, certain codes like the R-S codes can detect the presence of a catastrophic error by examining the received message.

The interleaver subsystem rearranges the encoded symbols over multiple code blocks. This effectively spreads out long burst noise sequences so they appear to the decoder as independent random symbol errors or shorter burst errors.

As it was shown in charter 4, (n, k) Reed Solomon codes are linear block non-binary cyclic codes with symbols, where k is the number of data symbols being encoded, and n is the total number of code symbols in the encoded block, $n-k=2t$ is the number of parity symbols. This encoded block of length n symbols is called a codeword. The number of errors that can be correct is $t=\frac{d_{\min}-1}{2}=\frac{n-k}{2}$.

Suppose there are number of p corrupted by errors symbols in the block. If $p \le t$ errors will be located and corrected. If bursts of length $p > t$ the error correcting code will fail. Let assume that $t < p \le t*i$, where i is called a interleaving depth (it is an integer):

$$i \ge p/t, \tag{52}$$

The Reed-Solomon (n,k) code can be used if we can spread the burst error sequence over several code blocks so that each block has no more than t errors which can then be corrected. This can be accomplished using block interleaving.

Example 4: There is RS (255,235) code with $t=10$, $i=5$. Without interleaver the data bytes output from the Reed-Solomon encoder would appear as shown below in Figure 5, where bytes numbered 0 to 234 are the data bytes and bytes 235 to 254 are the parity check bytes.

$a_0 \ldots a_{254}$	$b_0 \ldots b_{254}$	$c_0 \ldots c_{254}$	$d_0 \ldots d_{254}$	$e_0 \ldots e_{254}$

Figure 5. Data bytes output from the encoder for RS (255,235) code

The sequence of bytes enters in interleaver RAM row by row to build an $i*k$ data table as shown in Figure 6.

	Data bytes					Parity bytes			
codeword A	a_0	a_1	a_2	...	a_{234}	a_{235}	a_{236}	...	a_{254}
codeword B	b_0	b_1	b_2	...	b_{234}	b_{235}	b_{236}	...	b_{254}
codeword C	c_0	c_1	c_2	...	c_{234}	c_{235}	c_{236}	...	c_{254}
codeword D	d_0	d_1	d_2	...	d_{234}	d_{235}	d_{236}	...	d_{254}
codeword E	e_0	e_1	e_2	...	e_{234}	e_{235}	e_{236}	...	e_{254}

Figure 6. Interleaver Table – Data Bytes are Input Row by Row and Output Column by Column

To interleave the data bytes its is sent to the output devise column by column as it is shown in Figure 7.

a_0	b_0	c_0	d_0	e_0	a_1	b_1	...	a_{254}	b_{254}	c_{254}	d_{254}	e_{254}

Figure 7. Data Bytes Output from the Block Interleaver

6. Convolution code

Possibilities of convolutional coding have been introduced by Peter Elias (1923-2001) in 1955. Convolutional coding offers an alternative to block codes for transmission over a noisy channel (Convolutional encoding with Viterbi decoding is a Forward Error Correcting technique that is particularly suited to a channel in which the transmitted signal is corrupted mainly by additive white gaussian noise). It has been used in space communications and wireless communications. An advantage of convolutional coding is that it can be applied to a continuous data stream as well as to blocks of data.

Convolutional codes are commonly characterized by parameters: n, k, K, and r, where

n is a number of outputs bits;

k is a number of inputs bits;

K is a constraint length which equals the shift register stages of the encoder, i.e. the number of information bits in the register memory, which take a part in coding process;

r is a ratio of information bits to transmitted bits k/n, called a code rate; it is a measure of efficiency of the code. Smaller coding rate gives more powerful code due to extra redundancy and less bandwidth efficiency.

The principle of "overlap" is extensively used in convolutional codes. For example, in a rate k/n convolutional code, each input is "overlapped" with several previous inputs to produce each pair of encoded symbols. It is not able to divide coded sequence into blocks as each coded symbol pair is interlocked with its neighbours. Longer constraint lengths produce more powerful codes, but the complexity of decoding operations increases exponentially with constraint length.

6.1. Coder and coding.

A convolutional encoder is a Mealy[6] machine, where the output is a function of the current state and the current input. Figure 8 shows a (n,k,m) convolutional encoder.

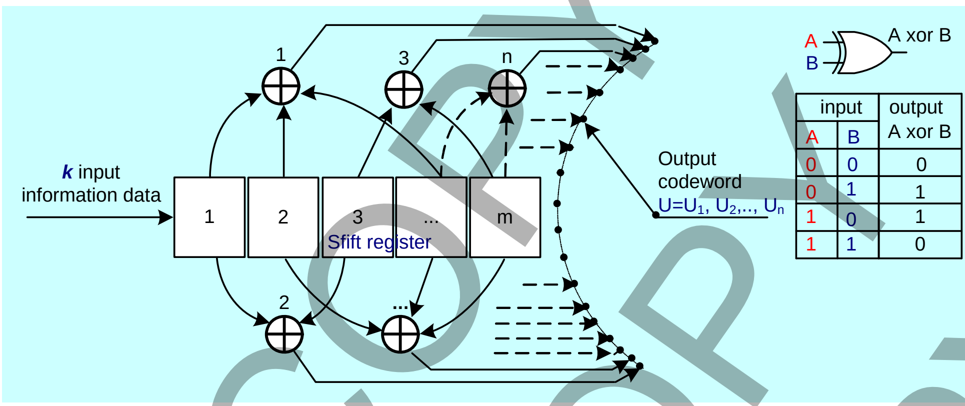


Figure 8. (n,k,m) convolutional encoder

It consists of one or more shift registers, multiple XOR gates and an output switch. The stream of information bits flows in to the shift register from one end and is shifted out at the other end every clock time. XOR gates are connected to some stages of the shift registers to generate the output. There is no theoretical basis for the optimal location of the shift register stages to be connected to XOR gates. It is based on an empirical approach. The location of stages as well as the number of memory elements determines the minimum Hamming distance. The minimum Hamming distance determines the maximal number of correctable bits.

The output switch consecutively reads the outputs of XOR gates and generates the output codeword $U_i = G(m_i)$. Frequency clock switching depends on code rate $r = k/n$. Thus, the frequency clock switching has to be n times higher by comparison with input clock frequency. In practical use there are systematic and non-systematic convolutional codes depending on the generator polynomial. For example, if the generator polynomials $g_1(x)=1$ or $g_2(x)=1$ for a rate 1/2 code, the information sequence would appear directly in the output and the code becomes systematic. Figure 9 shows idea of systematic convolutional code, figure 10 shows encoder of non-systematic convolutional code.

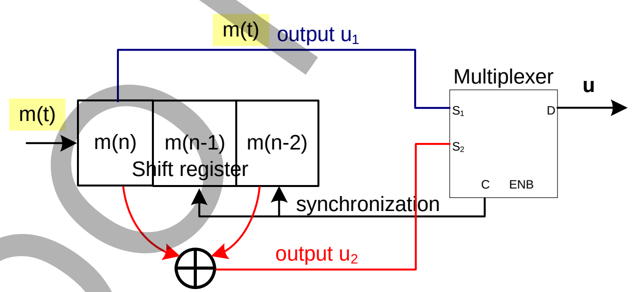


Figure 9. Coder of systematic convolutional code

[6] George H. Mealy „A Method for Synthesizing Sequential Circuits". 1955, *Bell Systems Technical Journal* 34: pp. 1045–1079.

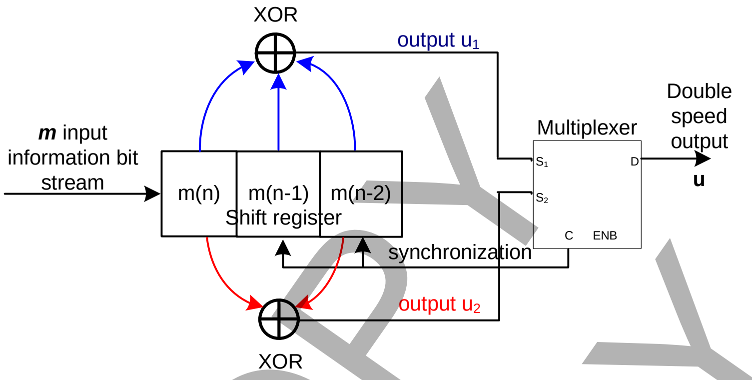


Figure 10. Coder of non-systematic convolutional code

One of the advantages of a systematic code is that it is simple to extract the information sequence for a decoder.

The encoder shown in figure 10 has a code rate $r = 1/2$. This means there are two output bits for each input bit. Here the output bits are transmitted one after another, two per clock cycle. The two output can be represent as binary vectors $[111]$ and $[101]$ are known as the generating vectors or generating polynomials for the code:

$$g_1(x) = 1 + x + x^2$$

$$g_2(x) = 1 + x^2 \; . \qquad (53)$$

where $m(n)$ is a last entered bit, $m(n-1)$ is a bit one clock time older than $m(n)$, and $m(n-2)$ is a bit two clock times older than $m(n)$.

After every clock time two bits appear in output: the first from output u_1 and the second from output u_2

$$u_1 = m(n) \oplus m(n-1) \oplus m(n-2)$$

$$u_2 = m(n) \oplus m(n-2) \; . \qquad (54)$$

The correlation encoder can be described as a Mealy machine. The state is the two bits in the shift register. Let the first input bit to the shift register be $m(n) = 1$, and let the flip-flops be reset to zero so $m(n-1) = 0$ and $m(n-2) = 0$. This state is

$$S_0 = 00 = [m(n-1), m(n-2)].$$

The output is

$$u_1 = m(n) \oplus m(n-1) \oplus m(n-2) = 1 \oplus 0 \oplus 0 = 1,$$

$$u_2 = m(n) \oplus m(n-2) = 1 \oplus 0 = 1,$$

$$u = [u_1, u_2] = 11$$

After the clock, state bit $m(n-1) = 0$ will shift right into $m(n-2)$, the input $m(n) = 1$ will shift right into $m(n-1)$, and the next state will be $S_1 = 10 = [m(n), m(n-1)]$. In the first stage can enter $m(n+1) = 0$ or $m(n+1) = 1$ with equal probability.

If there is $m(n+1) = 0$ the output is

$$u_1 = m(n+1) \oplus m(n) \oplus m(n-1) = 0 \oplus 1 \oplus 0 = 1,$$

$$u_2 = m(n+1) \oplus m(n-1) = 0 \oplus 0 = 0,$$

$u = [u_1, u_2] = 10$,

if there is $m(n+1) = 1$ the output is

$u_1 = m(n+1) \oplus m(n) \oplus m(n-1) = 1 \oplus 1 \oplus 0 = 0$,

$u_2 = m(n+1) \oplus m(n-1) = 1 \oplus 0 = 1$,

$u = [u_1, u_2] = 01$.

In a clock time 3 the bit $m(n-1) = 0$ out of register, the bit $m(n) = 1$ will be moved to right in the third register, the bit $m(n+1) = 0\ or\ 1$ will move in the second stage. In the first enters new bit $m(n+2)$, which can also be 0 or 1. The states in a clock 3 are $S_2 = 01 = [m(n+1), m(n)]$ or $S_3 = 11 = [m(n+1), m(n)]$. The output can be

If $m(n+2) = 0$, $m(n+1) = 0$:

$u_1 = m(n+2) \oplus m(n+1) \oplus m(n) = 0 \oplus 0 \oplus 1 = 1$,

$u_2 = m(n+2) \oplus m(n) = 0 \oplus 1 = 1$,

$u = [u_1, u_2] = 11$,

If $m(n+2) = 1$, $m(n+1) = 0$:

$u_1 = m(n+2) \oplus m(n+1) \oplus m(n) = 1 \oplus 0 \oplus 1 = 0$,

$u_2 = m(n+2) \oplus m(n) = 1 \oplus 1 = 0$,

$u = [u_1, u_2] = 00$.

If $m(n+2) = 0$, $m(n+1) = 1$:

$u_1 = m(n+2) \oplus m(n+1) \oplus m(n) = 0 \oplus 1 \oplus 1 = 0$,

$u_2 = m(n+2) \oplus m(n) = 0 \oplus 1 = 1$,

$u = [u_1, u_2] = 01$.

If $m(n+2) = 1$, $m(n+1) = 1$:

$u_1 = m(n+2) \oplus m(n+1) \oplus m(n) = 1 \oplus 1 \oplus 1 = 1$,

$u_2 = m(n+2) \oplus m(n) = 1 \oplus 1 = 0$,

$u = [u_1, u_2] = 10$.

Example 5: Let the information data consist three bits $m = [101]$. The encoder structure is shown in Fig. 10. The stages of the shift register are reset and the first bit received (clock time 1) is $m(n) = 1$. So in the first cycle in the output will

$u_1 = 1 \oplus 0 \oplus 0 = 1$, $u_2 = 1 \oplus 0 = 1$, $u = [u_1, u_2] = 11$.

After the clock a bit $m(n+1) = 0$ enters. The result is

$u_1 = 0 \oplus 1 \oplus 0 = 1$, $u_2 = 0 \oplus 0 = 0$, $u = [u_1, u_2] = 10$.

Then (at time 3) a bit $m(n+2) = 1$ enters. The result is

$u_1 = 1 \oplus 0 \oplus 1 = 0$, $u_2 = 1 \oplus 1 = 0$, $u = [u_1, u_2] = 00$.

After that a tail of two zero-bits is appended to data bits to clear out the memory after encoding the last bit (all-zero); the result is two more pairs of output symbols:

$t = 4$: $u = [u_1, u_2] = 10$,

$t=5$: $u=[u_1,u_2]=11$.
The output coded message is 11 10 00 10 11.

The ratio $R_{eff}=\dfrac{L}{n[(L/k)+(K-1)]}$ is called effective code rate, where L is the number of data bits. For example 5: $L=3$, $n=2$, $k=1$, $K=3$ and

$$R_{eff}=\frac{L}{n[(L/k)+(K-1)]}=\frac{3}{2[(3/1)+(3-1)]}=\frac{3}{10}$$ which is less than $r=1/2$.

The effective code rate falls below code rate r because the added zero-bits do not carry information.

Example 6: Let the information data consist three bits $m=[101]$. The encoder structure is shown in Fig. 11. This is a coder with a constraint length $K=3$ and code rate $r=1/3$.

The memory of the shift register is clean and the first bit received (clock time 1) is $m(n)=1$. So in the first cycle in the output will

$u_1=1\oplus 0\oplus 0=1$, $u_2=1\oplus 0=1$, $u_3=0\oplus 0=0$ $u=[u_1,u_2,u_3]=110$.

After the clock a bit $m(n+1)=0$ enters. The result is

$u_1=0\oplus 1\oplus 0=1$, $u_2=0\oplus 0=0$, $u_3=1\oplus 0=1$ $u=[u_1,u_2,u_3]=101$.

When $t=3$ $m(n+2)=1$ enters. The output result is

$u_1=1\oplus 0\oplus 1=0$, $u_2=1\oplus 1=0$, $u_3=0\oplus 1=1$ $u=[u_1,u_2,u_3]=001$.

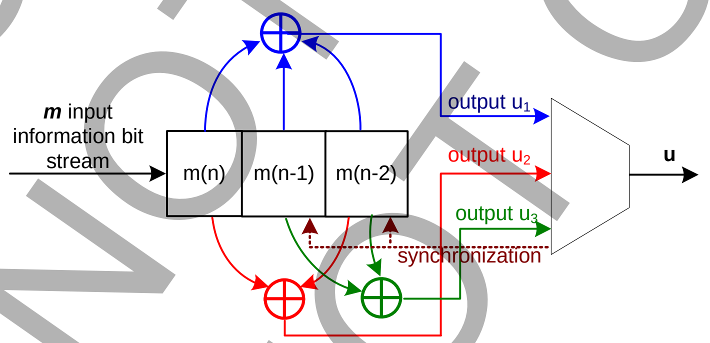


Figure 11. Coder with a constraint length $K=3$ and code rate $r=1/3$

After that a tail of two zero-bits is appended to data bits to clear out the memory after encoding the last bit (all-zero).

$t=4$: $u_1=0\oplus 1\oplus 0=1$, $u_2=0\oplus 0=0$, $u_3=1\oplus 0=1$ $u=[u_1,u_2,u_3]=101$.

$t=5$: $u_1=0\oplus 0\oplus 1=1$, $u_2=0\oplus 1=1$, $u_3=0\oplus 1=1$ $u=[u_1,u_2,u_3]=111$.

The output coded message is 110 101 001 101 111.

The coder is represented by set of generator polynomials, one for each of the n modulo-2 adders. Each polynomial is of degree $K-1$ or less and describes the connection of the shift register to that modulo-2 adder. The coefficients can be either 1 or 0. The value 1 means that connection exists between the stage of shift register and the modulo-2 adder. The value 0 means that connection does not exist. For instant, the encoder shown in figure 10 can be representing by generator polynomials (53), the encoder shown in figure 11 can be representing by generator polynomials

$$g_1(x) = 1 + x + x^2$$

$$g_2(x) = 1 + x^2$$

$$g_3(x) = x + x^2 . \qquad (55)$$

The information message from Example 5 can be representing by generator polynomial $m(x) = 1 + x^2$. The output u_1 is formed by multiplication of $g_1(x)$ with $m(x)$:

$$g_1(x).m(x) = (1 + x + x^2).(1 + x^2) = 1 + x + x^3 + x^4,$$

and

$$g_2(x).m(x) = (1 + x^2).(1 + x^2) = 1 + x^4 .$$

These Equations can represent with coefficients:

$$g_1(x).m(x) = 1 + 1.x + 0.x^2 + 1.x^3 + 1.x^4 ;$$

$$g_2(x).m(x) = 1 + 0.x + 0.x^2 + 0.x^3 + 1.x^4 ;$$

$$U(x) = (1,1) + (1,0).x + (0,0).x^2 + (1,0).x^3 + (1,1).x^4 ;$$

$$U = G(m) = 11 \quad 10 \quad 00 \quad 10 \quad 11 .$$

The result of $U(x)$ is the same as that in Example 5.

For Example 6:

$$g_1(x).m(x) = (1 + x + x^2).(1 + x^2) = 1 + 1.x + 0.x^2 + 1.x^3 + 1.x^4$$

$$g_2(x).m(x) = (1 + x^2).(1 + x^2) = 1 + 0.x + 0.x^2 + 0.x^3 + 1.x^4$$

$$g_3(x).m(x) = (x + x^2).(1 + x^2) = 0 + 1.x + 1.x^2 + 1.x^3 + 1.x^4$$

$$U = 110\ 101\ 001\ 101\ 111 .$$

Example 7: Let the information data consist three bits $m = [1011]$. The encoder structure is shown in Fig. 10. This is a coder with a constraint length $K = 3$ and code rate $r = 1/2$.

The work of the encoder to create a codeword can be presented in tabular form, as it is shown below.

№ of cicle	Input bit, cycle i	Contents of the register	State, cycle i	State, cycle $i+1$	u_1	u_2	u
		000	$S_0 = 00$	$S_0 = 00$	-	-	-
1	1	100	$S_0 = 00$	$S_1 = 10$	1	1	11
2	0	010	$S_1 = 10$	$S_2 = 01$	1	0	10
3	1	101	$S_2 = 01$	$S_1 = 10$	0	0	00
4	1	110	$S_1 = 10$	$S_3 = 11$	0	1	01
5	0	011	$S_3 = 11$	$S_2 = 01$	0	1	01
6	0	001	$S_2 = 01$	$S_0 = 00$	1	1	11

The output codeword is $11\,10\,00\,01\,01\,11$.

6.2. State representation and state diagram

A convolutional encoder is a Mealy machine (finite-state machine), where the output is a function of the current state and the current input. There are a finite number of states that the machine can encounter. The state consists of the smallest quantity of information that with a current input bit can predict the output of the machine. For a rate $1/n$ convolutional encoder the state is represented by contents of $(K-1)$ rightmost stages.

There are some alternative ways of describing a convolutional code. It can be expressed as a state diagram, a tree diagram, or a trellis diagram.

A State Diagram shows all possible present states of the encoder as well all the possible state transitions that may occur. In order to create the state diagram, a state transition table may first be made, showing the next state for each possible combination of the present state and input to the decoder. For example it can draw the table for the encoder shown in figure 10.

Table 4: State Transition Table

Current State	Next State if	
	input $m=0$	input $m=1$
00	00	10
10	01	11
01	00	10
11	01	11

Table 5 shows the change in output for each combination of input and previous output.

Table 5: Output Table

Current output	Output simbols if	
	input $m=0$	input $m=1$
00	00	11
10	10	01
01	11	00
11	01	10

The state diagram is created using the information from Table 4 and Table 5, as shown in Figure 12.

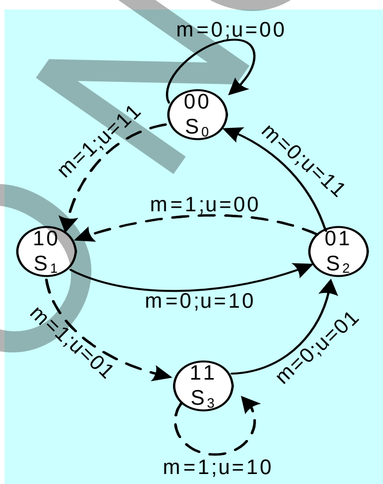


Figure 12. State Diagram (coder K=3, $R=1/2$, Figure 10)

For each state, there can be two outgoing transitions; one corresponding to a '0' input bit and the other corresponding to a '1' input bit. The values inside the circles indicate the state (contents of $(K-1)$ rightmost stages). The values on the arrows indicate the output of the encoder. A solid line corresponds to input 0, a dotted line – to input 1. The State Diagram shows an entire picture the all possible states and their output code words. The disadvantage of the state diagram is the lack of measurement of time, i.e. it is not clear what state of the encoder in which time is obtained.

A Tree Diagram shows the passage of time. In each cycle coding procedure is described in a passage in the branches from left to right as each branch shows the codeword. If the input bit is 1, the codeword is found by moving to the next right and below branch, if the input bit is zero - the right and above. Any input sequence can be traced through a path in this diagram which forms the corresponding code word. This path can also be called the code word path. For instance, Figure 13 shows a tree diagram of coder, shown in Figure 10.

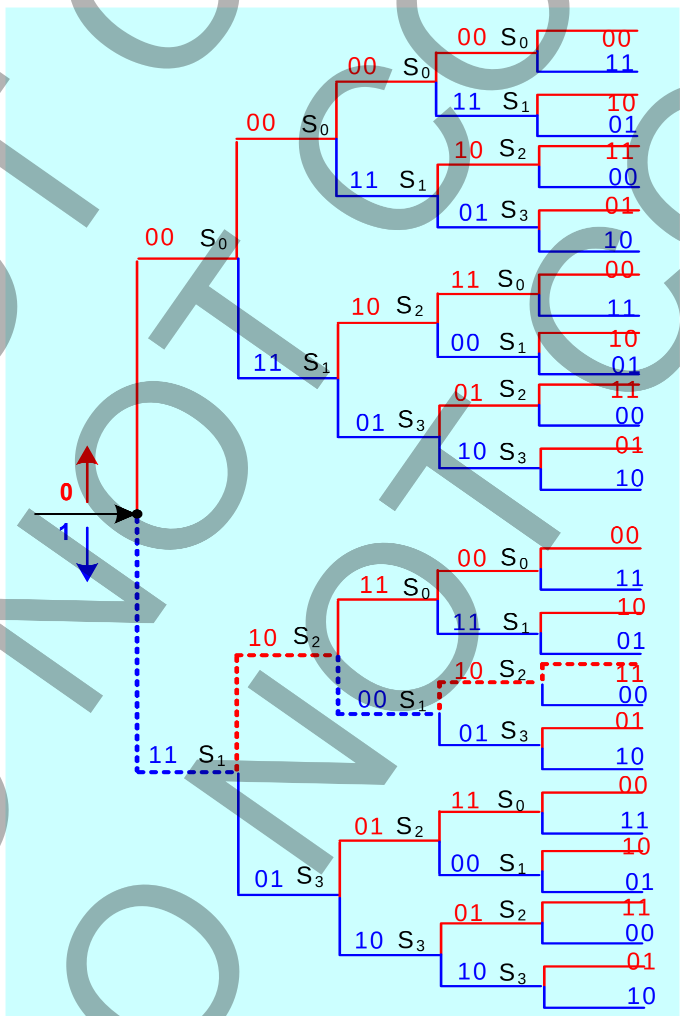


Figure 13. Tree Diagram of coder, shown in Figure 10

The disadvantage is the progressive increase in the branches if information message bits (input) are too much. After the first K branches (constraint length) tree is repeated.

Trellis Encoding Diagram present the sequence of events in linear time in a compact form. In constructing the chart along the horizontal axis has time clocks and the vertical - all possible states. Figure 14 shows first two clocks of a trellis diagram for the encoder given in Figure 10.

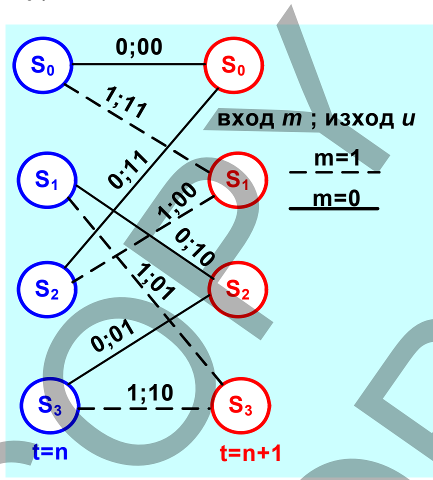


Figure 14. Trellis structure for R=1/2, k=3 convolutional code (first two clocks)

In the trellis diagram, nodes correspond to the states of the encoder. From an initial state (S_0) the trellis records the possible transitions to the next states for each possible input pattern. Figure 15 shows a trellis diagram for the encoder given in Figure 10 and input message $m = 101100$.

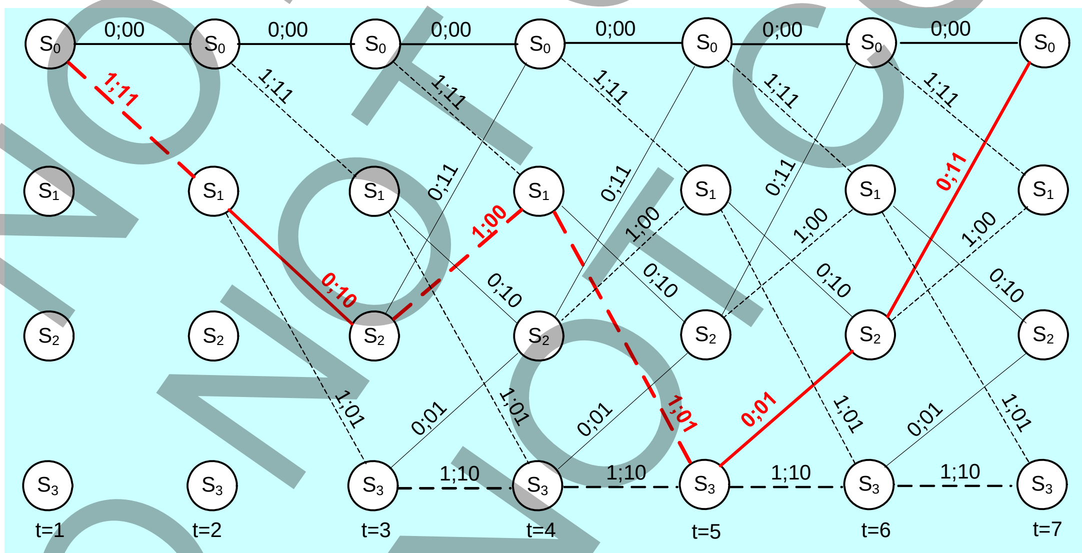


Figure 15. Trellis Diagram for the encoder given in Figure 10 and input message $m = 101100$

For this code, there are four possible encoder states. Each row of nodes is representing the same state of the encoder at different time steps. The first row nodes shows the state $S_0 = 00$, the second - $S_1 = 10$, the third - $S_2 = 01$, and fifth - $S_3 = 11$. Thus, the trellis diagram requires 2^{K-1} nodes to represent 2^{K-1} possible encoder states. Each of the states can be entered from either of two preceding states. At every time step there are two branches output from each node where an input “0” to the encoder corresponds to the upper branch and a “1” input to the lower branch. Since the initial condition of the encoder is $S_0 = 00$, and the two memory flushing bits are zeroes, the lines start out at $S_0 = 00$ and end up at the same state. For example, the input sequence

1 0 1 1 0 0 corresponds to the particular path, shown with thick red lines, through the trellis. The output codeword is $11\,10\,00\,01\,01\,11$, and states are $S_0, S_1, S_2, S_1, S_3, S_2, S_0$.

6.3. The Viterbi Decoding Algorithm

There are two types of decoding algorithms used with convolutional encoding: sequential decoding and Viterbi decoding algorithm. Sequential decoding has the advantage that it can perform very well with long-constraint-length convolutional codes, but it has a variable decoding time. The Viterbi decoding algorithm was developed by Andrew J. Viterbi[7] in 1967. This algorithm performs a maximum likelihood decoding (maximum likelihood detector). Viterbi decoding has the advantage that it has a fixed decoding time and it has good performance in hardware. Disadvantage is exponential growth of its computational requirements as a function of the constraint length. In this reason, it is usually limited to constraint lengths of $K = 9$ or less.

The algorithm includes calculating a distance between received signal at time $t_i, i = 1,2,3,...$ and all trellis paths entering each state at time t_i. The metric of calculating is the Hamming distance between the received channel symbol pair and the all possible channel symbol pairs. The Hamming distance is computed by counting how many bits are different between the received channel symbol pair and the possible channel symbol pairs (see part 3). The results can be zero, one, ... The Hamming distance values is computing at each time instant for the paths between the states at the previous time instant and the states at the current time instant are called branch metrics. When the encoding process is presented with trellis diagram it shows that the coded output word is formed as a path through the states of the encoder at any clock time linear in time. The decoder receives the sequence of bits of the codeword and must determine the path through the states that will determine the decoded sequence of bits. Since it is unknown whether the received sequence without error, the path in trellis diagram is chosen according to the smallest Hamming distance.

Figure 16 shows how to choice the path in trellis diagram according to the smallest Hamming distance measured in every clock time. Suppose the received bits have an error so instead of $11\,10$, one receives $11\,1\underline{1}$. The lines start out at $S_0 = 00$ and according to two possible outputs (00 and 11) the path can only go through state $S_0 = 00$ or state $S_1 = 10$. Suppose at $t = 1$, it is received 11. The only possible channel symbol pairs that could have received are 00 and 11. The Hamming distance between 11 and 11 is zero. The Hamming distance between 11 and 00 is two. The Hamming distance between the received input and the bits for the transition is shown in box. The choice is $S_0 \to S_1$ because of less Hamming distance. The next input has an error. Possible paths, 10 or 01, are at Hamming distance 1. At $t = 2$ it can going from S_3 to state S_3 (Hamming distance is 1) or from S_3 to state S_2 (Hamming distance is 1, too), or from S_2 to state S_1 (Hamming distance is 0) and from S_2 to state S_0 (Hamming distance is 2). Thus at $t = 1$ the proper path was not obvious, at $t = 2$, the choice is clearer. The choice is a path through the trellis based on the path Hamming distance or path metric,

[7] Viterbi A. J., Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm", *IEEE Transactions on Information Theory*, Volume IT-13, pages 260-269, April, 1967.

which is the sum of the Hamming distances as one steps along a path through the trellis. The path $S_0 \to S_1 \to S_2 \to S_1$ has a Hamming distances sum equal to 1. The Hamming distances sums of other paths are higher then this one. At $t = 3$ it can going from S_1 to state S_3 (Hamming distance is 0) or to S_2 (Hamming distance is 2). Thus, the choice is going according to minimum Hamming distance in every clock time.

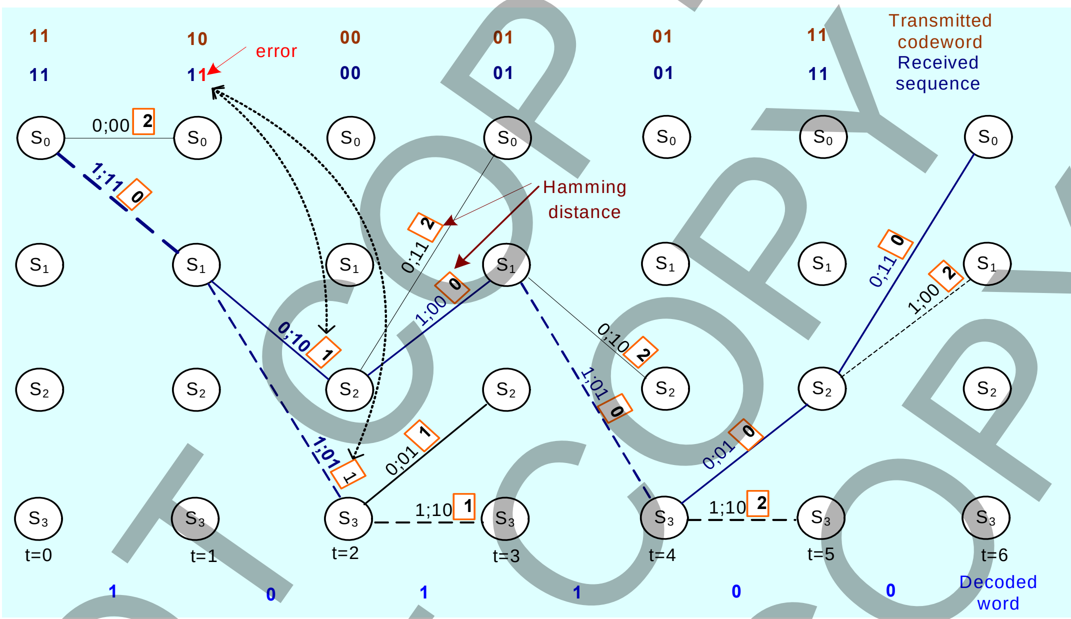


Figure 16. To choice the path in trellis diagram according to the smallest Hamming distance

Figure 17 illustrates how to choice the path in trellis diagram according to the smallest sum of the Hamming distances that is called path metric.

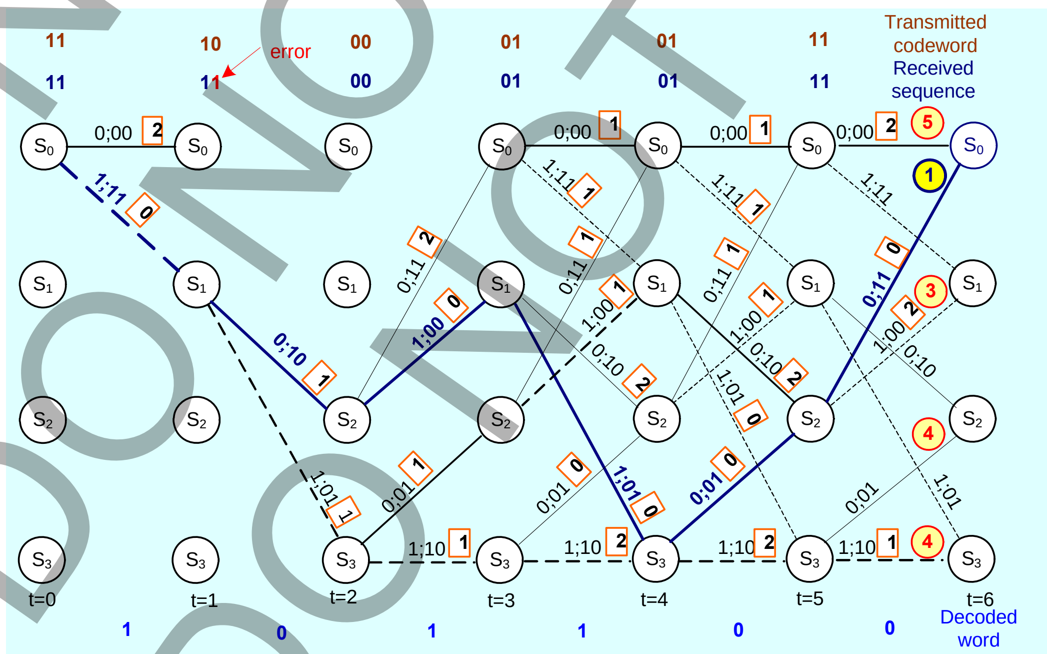


Figure 17. The choice the path according to the smallest sum of the Hamming distances (path metric)

At $t = 6$, one path has a total distance of 1 from the input data. The others have a distance of 3, 4 or 5. Thus the most likely path is $S_0, S_1, S_2, S_1, S_3, S_2, S_0$ with a path distance of 1, and the corresponding output decoded word is 101100. In decoders where the input is an analog signal, the distance between the actual and expected voltage may be measured, and the sum of the squares of the errors might be used for the branch metric.

The minimal path metric is the rule that the Viterbi decoder exploits to recover the original message. It is the maximum-likelihood path.

6.4. Implementation of the Viterbi Decoder

A Viterbi decoding algorithm consists of the three major parts, shown in Figure 18:

1. Branch metric calculation – calculation of a distance between the input pair of bits and the other possible pairs (“00”, “01”, “10”, “11”).

2. Path metric calculation – for every encoder state, calculate a metric for the survivor path (a path with the minimum metric) ending in this state.

3. Traceback (survivor path decoding) – it is necessary for hardware implementations that don't store full information about the survivor paths, but store only one bit decision every time when one survivor path is selected from the two.

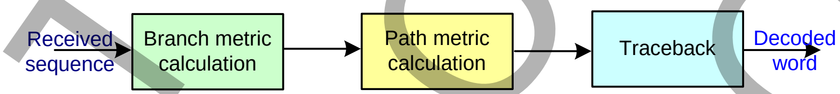


Figure 18. The major parts of the Viterbi decoding algorithm

The Branch metric unit (BMU) receives input data from the channel and computes a metric for each state and input combination. Branch metric calculation is different for hard decision and soft decision decoders. For a hard decision decoder, a branch metric is a Hamming distance between the received pair of bits and the “ideal” pair. Therefore, for every input pair a branch metric can take values of 0, 1 and 2. For a soft decision decoder, the received code words are quantised into different levels according to the signal strength then the BMU maps the levels of code words into BMs according to their likelihood. For a soft decision decoder, a branch metric is measured using the Euclidean distance:

$$B_m = (x - x_0)^2 + (y - y_0)^2, \tag{56}$$

where x is the first received bit in the pair, y – the second, x_0 and y_0 – the “ideal” values.

Path metric is calculated by a number of units called “Add-Compare-Select” (ACS). The ACS unit adds the current metric to the accumulated metric for each path and determines the least metric for each state of the trellis. The ACS retrieves the accumulated metric from the register files, and then adds the current metric. The result is stored back in the register files. This procedure is repeated for every encoder state.

- Add – a new value of the state metrics has to be computed at each time instant; for a given state are known two states on the previous step which can move to this state, and the output bit pairs that correspond to these transitions. To calculate new path metrics, it has to add the previous path metrics with the corresponding branch metrics, i.e. the state metrics have to be updated every clock cycle.

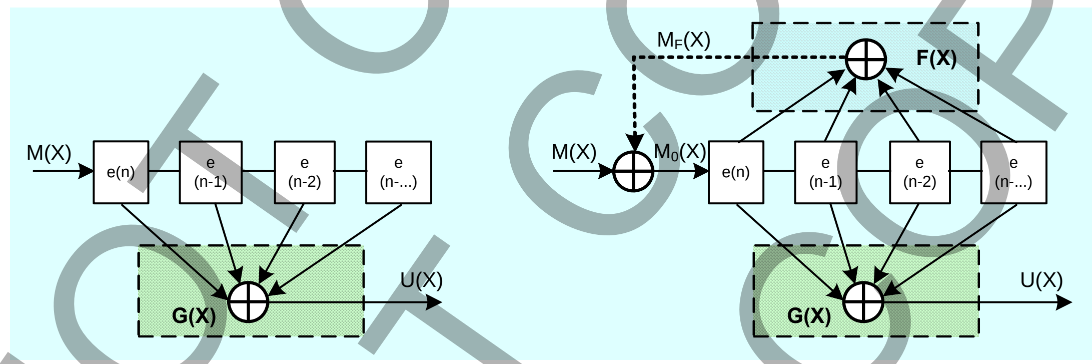
- Compare and select – There are two paths, ending in a given state. One of them (with greater metric) is rejected.

As there are 2^{K-1} encoder states, we have 2^{K-1} survivor paths at any given time.

The traceback unit (TBU) traces back the trellis after a block of data (determined by the trace back length) has been processed by the ACS. First, the TBU establishes an optimal path by starting from the node of minimum metric and traces back the path in the trellis all the way to the beginning of the trellis diagram. Then, the original data is determined.

6.5. Recursive Systematic Convolutional Encoder

The recursive systematic convolutional (RSC) encoder is produced from the nonrecursive nonsystematic (conventional) convolutional encoder by feeding back one of its encoded outputs to its input. Figure 19a,b shows an example of conventional convolutional encoder and recursive systematic convolutional encoder.



a) b)

Figure 19. Example of conventional convolutional encoder and recursive systematic convolutional encoder.

These encoders are represented by: $M(X)$ is a polynomial of information message; $U(X)$ is a coded message (word); $G(X)$ is a nonrecursive generator polynomial; $F(X)$ is a feedback polynomial. The output transfer function for the encoder, shown in Figure 19a, is

$$U(X) = M(X).G(X), \tag{57}$$

for the recursive encoder

$$G_R(X) = U(X)/M(X), \tag{58}$$

$$M_0(X) = M(X) + M_F(X),\ M_F(X) = M_0(X).F(X),$$

$$M_0(X) = M(X) + M_0(X).F(X) \to M(X) = M_0(X)[1-F(X)],\ U(X) = M_0(X).G(X)$$

From Equations (57) and (58) the recursive polynomial is

$$G_R(X) = U(X)/M(X) = G(X)/[1-F(X)], \tag{59}$$

i.e. the recursive polynomial depends on the nonrecursive generator polynomial $G(X)$ and the feedback polynomial $F(X)$.

It has been proven that the feedback polynomial $F(X)$ has to be a primitive polynomial as a condition for stable operation of the recursive encoder.

Advantages of the recursive systematic convolutional encoder (the positive effect of the feedback) are:

- Reducing the number of states of the encoder, which increases the efficiency of decoding algorithm, for example, if the degree of the nonrecursive generator polynomial $G(X)$ is p, the degree of the feedback polynomial $F(X)$ is q and $q < p$, after division the degree of the recursive polynomial decreases, which reduces the constraint length of recursive code; this allows for synthesis of recursive code with a small number of states and simplified decoding algorithm;

- Increasing the average distance between output code sequences (increasing the code Hemming weight), which achieves better results for error correction.

References

1. Bernard Sklar, Digital Communications: Fundamentals and Applications, Second Edition, Prentice-Hall, 2001.
2. Berrou C., Glavieux A., Thitimajshima P., Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes // Proc. of the Intern. Conf. on Commun (Geneva, Switzerland). May 1993, pp.1064–1070.
3. Berrou C., Glavieux A., Near Optimum Error Correcting Coding and Decoding: Turbo-Codes, IEEE Trans. On Comm., Vol. 44, No. 10, October 1996.
4. Blahut, R. E., Theory and Practice of Error Control Codes, MA: Addison-Wesley, 1983.
5. C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech., 1948.
6. Dennis Roddy, Satellite communications, McGraw-Hill Professional, 2001 - 569 p.
7. Fred Daneshgaran, Massimiliano Laddomada, and Marina Mondin, High-Rate Recursive Convolutional Codes for Concatenated Channel Codes, IEEE Transactions on Communications, Vol. 52, No. 11, november 2004.
8. Fred Ma, John Knight, Convolution Codes, http://read.pudn.com/, 2001.
9. Forney G., D., The Viterby algorithm. IEEE, vol. 61, n. 3, Marsh, 1978, pp. 268-278.
10. Gallager, R. G., Information Theory and Reliable Communication, New York: John Wiley and Sons, 1968.
11. Gérard Maral, Michel Bousquet, Satellite communications systems: systems, techniques, and technology, John Wiley and Sons, 2002 - 757 pp.
12. George H. Mealy „A Method for Synthesizing Sequential Circuits". 1955, *Bell Systems Technical Journal* 34: pp. 1045–1079.
13. Imai H., Hirakawa S., A New Multilevel Coding Method Using Error Correcting Codes. IEEE Trans. Info Theory, vol. IT-23, n. 3, pp. 371–377, May 1977.
14. Intuitive Guide to Principles of Communications, Coding Concepts and Block Coding, www.complextoreal.com
15. Interleaving for Burst Error Correction, AHA products groop, Moscow, Idaho, USA, www.aha.com.
16. International Telecommunications Union, www.itu.int.
17. Ippolito, Louis J., Satellite communications systems engineering: atmospheric effects, satellite link design, and system Performance, JohnWiley & Sons Ltd, 2008, 396 p.
18. Mealy, George H., A Method for Synthesizing Sequential Circuits, Bell Systems Technical Journal 34,1955: pp. 1045–1079.
19. Michael Kolawole, Satellite communication engineering, Marcel Dekker, 2002, 263 p.
20. RPC Telecommunications Ltd., http://www.satcom.co.uk/.
21. Soleymani M. R., Yingzi Gao, Vilaipornsawai U., Turbo coding for satellite and wireless communications. Kluwer Academic Publishers, Dordrecht. 2002. 214 pp.

22. Stewart Robert W., Daniel Garsia-Alis, Concise DSP Tutorial from Digital Communications: Fundamentals and Applications, Blue Box Multimedia, 2001, http://www.DSPedia.com

23. Viterbi A. J., Convolutional codes and their performance in communication systems, IEEE Trans. Commun., vol. 10, pp. 751–772, Oct. 1971.

24. Viterbi A. J., Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm", IEEE Transactions on Information Theory, Volume IT-13, pages 260-269, April, 1967.

Satellite Communications Systems

Lecture 4.
Multiple Access in Satellite Communications

1. Introduction of Multiple Access in Satellite Communications
2. Frequency Division Multiple Access
2.1. Frequency Division Multiple Access with PCM/TDM/PSK application
2.2. Frequency Division Multiple Access with PCM/SCPC/PSK application
3. Time Division Multiple Access
3.1. Time Division Multiple Access with PCM/TDM/PSK application
3.2. TDMA Frame Efficiency
3.3. TDMA capacity
3.4. Switching in satellite TDMA
4. Code Division Multiple Access
4.1. Direct Sequence Spread Spectrum
4.2. Frequency Hopping Spread Spectrum.

1. Introduction of Multiple Access in Satellite Communications

A multi-user system achieves a specific point in the capacity region depending on how the multi-user channel is shared by the users, which depends on the multiple access technique or Multiple Access (MA). It is the last technique before signal transmitting.

Satellite links are designed to provide desired link availability for average conditions. Satellite Multiple Access techniques interconnect ground stations through multiple satellite transponders. The goal is optimizing some of system parameters such as power efficiency; spectral efficiency and etc. These techniques are applicable to both fixed and mobile users. The satellite transponder may have different configurations, depending on the application and the satellite payload design.

Figure 1 illustrates four basic multiple access configurations[1].

The first configuration is single channel per carrier and single carrier per transponder (a). Single baseband channel modulate an RF carrier. The baseband channel could be analog (voice or video) or a digital (data, voice, or video).The modulation could be analog (amplitude or frequency modulation), or digital (frequency shift keying or various forms of phase shift keying). The second possibility is multiple channels per carrier and multiple carriers per transponder (b). To avoid intermodulation noise, the final amplifier in the transponder is usually operated in a power backoff mode. The next case (c) shows that multiple baseband channels are multiplexed onto a single data stream before carrier modulation (multiple channels per carrier, single carrier per transponder). The last case consists

[1] Ippolito, Louis J., Satellite communications systems engineering : atmospheric effects, satellite link design, and system Performance, 2008 JohnWiley & Sons Ltd, 396 pp.

of multiple multiplexed baseband channels modulating multiple RF carriers, with the multiple carriers per single transponder (d). This case also requires operating in a power backoff mode to avoid intermodulation noise. Cases (a) and (c) is called single carrier per transponder. The final amplifier can operate in full power, and provide high power efficiency. When configuration is multiple carriers per transponder (b and d) power backoff can be several dB, resulting in lower power efficiency in comparison with cases (a) and (c).

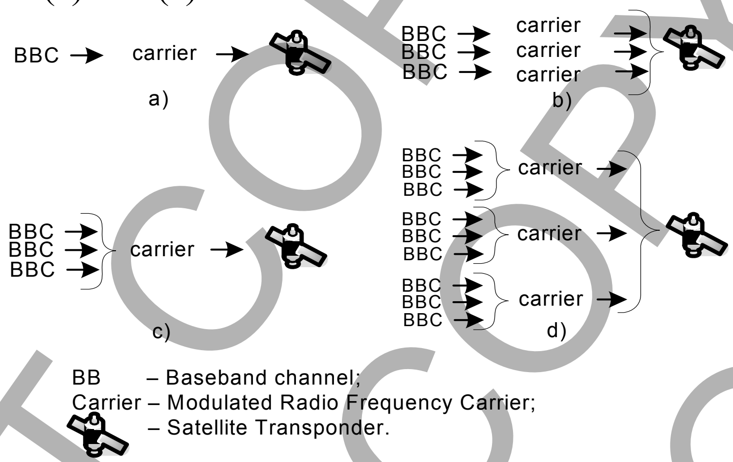


Figure 1. Illustration of the four basic multiple access configurations

The satellite multiple access methods can be categorized into three mainly three conventional techniques:

- Frequency Division Multiple Access (FDMA);
- Time Division Multiple Access (TDMA);
- Code Division Multiple Access (CDMA).

FDMA system separates multiple carriers in the transponder by *frequency*. The transmissions can be analog or digital, or combinations of both. TDMA system separates the multiple carriers by *time* in the transponder, presenting only one carrier at any time to the transponder. TDMA is most practical for digital data only, because the transmissions are in a burst mode to provide the time division capability. CDMA is a *combination of both frequency and time separation*. As the most complex technique, CDMA requires several levels of synchronization at both the transmission and reception levels and it is implemented for digital data only. CDMA offers the highest power and spectral efficiency operation of all three techniques.

On the base of these three fundamental techniques are developed some secondary access techniques which are putting into practice:

✓ Demand Assigned Multiple Access (DAMA) - changes signal configuration to respond to changes in user demand. FDMA or TDMA networks can be operated as an assigned-on-demand DAMA network. CDMA is a DAMA network by design, because it is a random access system by its implementation.

✓ Space Division Multiple Access (SDMA) - is a promising multiple access technology for improving capacity by the spatial filtering capability of adaptive antennas. It spatially separated physical links by different antenna beams, cells, sectored antennas, signal polarization, etc. It is employable with any of the three basic multiple access techniques. SDMA is an essential element of mobile satellite networks, which employ multibeam satellites and it can be applied to all other multiple access schemes.

✓ Satellite Switched TDMA (SS/TDMA) - employs sequenced beam switching to add an additional level of multiple access in a frequency translation satellite. The switching is accomplished at radio frequency (RF) or at an intermediate frequency (IF).

✓ Orthogonal frequency division division multiple access (OFDMA) - a special form of multicarrier modulation, can be used for multiplexing for multiple users. An OFDMA system is defined as one in which each user occupies a subset of subcarriers, and each carrier is assigned exclusively to only one user at any time. Advantages of OFDMA over OFDM-TDMA and OFDM-CDMA include elimination of intracell interference and exploitation of network/multiuser diversity.

2. Frequency Division Multiple Access

Frequency division multiple access (FDMA) was the first technique implemented on satellite systems. FDMA divides all frequency band into many frequency channels and assigns a separate frequency channel on demand to each earth station. Figure 2 shows functions of FDMA process.

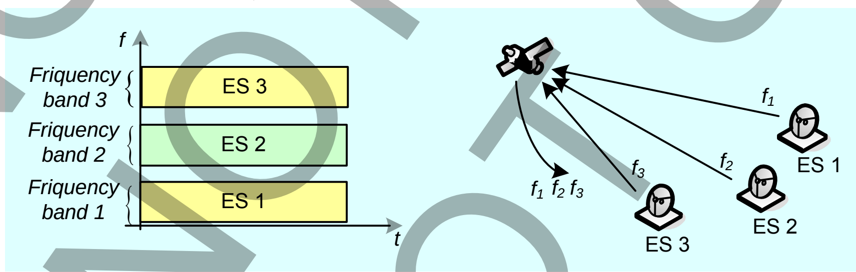


Figure 2. Principle of the Frequency Division Multiple Access

Figure 2 shows an example for three earth (ground) stations (ES) accessing a single frequency translation satellite transponder. Each station is assigned a specific frequency band for its uplink (f_1, f_2 *and* f_3). Frequency guard bands are used to avoid interference between the user slots. All stations receive total spectrum on downlink. The receiving station must be able to receive the full spectrum and can select the desired carrier for demodulation or detection.

The main advantages of this technique are:

- independent of the channels;
- no framing or synchronization bits needed;
- simplicity;
- transmissions can be analog or digital, or combinations of both;
- FDMA is most useful for applications where a full time channel is desired – for example, video distribution;

- the least expensive to implement technique.

The main disadvantages are:

- lower power efficiency because of the multiple carriers in the transponder final power amplifier requires power backoff operation to avoid intermodulation noise;

- lower flexibility if it is necessary another system configuration (change frequency plan);

- when number of users increase system efficiency decrease.

Multiple access system performance must be analyzed by considered the specific signal processing used in the satellite communications for both analog and digital signals.

2.1. Frequency Division Multiple Access with PCM/TDM/PSK application.

Some of the most common techniques in satellite communications are a PCM/TDM (pulse code modulation/time division multiplexed). Its application used for voice communications. By PSM analog signals are converted into digital and combined using TDM hierarchy according ITU-R Recommendation. The first level consists of 30 channels which of them with $64 kbit/s$, multiplexed to a $2{,}048 Mbit/s$ TDM bit rate. The carrier modulation is phase shift keying (BPSK or QPSK).

One of the most important parameter for evaluation is the capacity of the multiple access system. It determines the maximum number of users that can access the satellite and serves as the basis for decisions on demand access options on the link. The capacity for the PCM/TDM/PSK/FDMA digital multiple channel per carrier (MCPC) system is determined by the following steps[2].

1) Determine the composite carrier-to-noise density available on the RF link $(C/N_0)_T$.

2) Determine the required carrier-to-noise ratio $(C/N)_t$ required to support each individual MCPC carrier at the desired BER

$$\frac{C}{N} = \frac{E_b}{N_0} . \frac{R_b}{W_N}, \quad \left(\frac{C}{N}\right)_t = \left(\frac{E_b}{N_0}\right)_t + R_b - W_N + M_i + M_A, \ dB, \tag{1}$$

where R_b is a data rate of digital signal, in dB; W_N - noise bandwidth of carrier, in dB; C - carrier power; $\left(\frac{E_b}{N_0}\right)_t$ - the ratio, required for the threshold BER; M_i - MODEM implementation margin, in dB ($\approx (1 \div 3) dB$); M_A - Adjacent Channel Interference margin, in dB $\approx (1 \div 2) dB$.

The bit rate required to support each channel depends on the specific PCM baseband formatting: PCM: 64 kbps/voice channel, ADPCM: 32 kbps/voice channel.

The noise bandwidth depends on the carrier modulation

---

[2] Ippolito, Louis J., Satellite communications systems engineering : atmospheric effects, satellite link design, and system Performance, 2008 JohnWiley & Sons Ltd, 396 pp.

for BPSK $W_N = (1{,}2 * TDM\ bit\ rate) + 20\%$,

for QPSK $W_N = (1{,}2 * \frac{TDM\ bit\ rate}{2}) + 20\%$. (2)

The coefficient 1,2 account for differences between real and ideal characteristics of bandpass filter and the 20% factor is included to account for guard bands.

For example, the first level of hierarchy structure for $64 kbit/s$ PCM voice, TDM bit rate is $2{,}048 Mbit/s$, then

BPSK $W_N = (1{,}2 * 2048) + 0{,}2(1{,}2 * 2048) = 2{,}949 MHz$,

QPSK $W_N = (1{,}2 * \frac{2048}{2}) + 0{,}2(1{,}2 * \frac{2048}{2}) = 1{,}474 MHz$.

3) Determine the required carrier-to-noise *density* for each carrier:

$$\left(\frac{C}{N_0}\right)_t = \left(\frac{C}{N}\right)_t + W_N, \quad dB \tag{3}$$

4) Compare $(C/N_0)_t$ with the total available RF link carrier-to-noise ratio $(C/N_0)_T$ from 1) to determine the number of carriers, n_p, that can be supported:

$$\left(\frac{C}{N_0}\right)_T = \left(\frac{C}{N_0}\right)_t + 10\log n_p \tag{4}$$

From equation (4) it can be determinate the number of carriers, which can be supported:

$$n_p = 10^{\frac{\left(\frac{C}{N_0}\right)_T - \left(\frac{C}{N_0}\right)_t}{10}} \tag{5}$$

This result (5) has to round to the next lowest integer. It gives the power limited capacity of the system.

5) Determine the bandwidth limited capacity of the system, n_W, from

$$n_W = \frac{w_{TR}}{w_N}, \tag{6}$$

where w_{TR} is a satellite transponder bandwidth, w_N - noise bandwidth (including guard bands).

The result (6) (rounded to the next lowest integer), gives the *bandwidth limited capacity* of the system.

6) Determine the system capacity, ***C***, from the lower of the values of n_p or n_W:

$$C = n_p \ or \ n_B \tag{7}$$

The result ***C*** from (7) is the maximum number of carriers that can be used in the PCM/TDM/PSK/FDMA link, within the power and bandwidth limitations of the system.

2.2. Frequency Division Multiple Access with PCM/SCPC/PSK application.

PCM/SCPC/PSK is a digital baseband single carrier per channel (SCPC) system used for data and voice applications. It is not necessity of multiplexing. Each incoming signal is analog to digital converted, and modulated RF carrier by BPSK or QPSK for transmission over the satellite channel. A pair of channel frequencies is used for voice communications, one for each direction of transmission.

One advantage of this SCPC FDMA approach is that it can operate as a demand assignment access, where the carrier is turned off when not in use. The system can also use *voice activation*, which makes use of the statistics of voice conversations to share the SCPC carrier with multiple users.

It is known, that a typical voice channel conversation is active only about 40% of the time in any one direction. A *voice activation factor* (VA) is used to quantify the improvement possible in the network. For example, a 36MHz transponder has a bandwidth limited capacity of 800 SCPC channels, using 45 KHz channel spacing: $36.10^3/45 = 800$. The 800 channels correspond to 400 simultaneous conversations.

The capacity calculations for PCM/SCPC/PSK/FDMA are similar to the procedure, which is discussed in the previous section 1.1. Single channel carrier noise bandwidths and data rates are used to determine performance. The VA factor, if used, is applied to increase the power limited capacity, n_p, (4)). The bandwidth limited capacity, n_W, (5) is determined from

$$n_B = \frac{w_{TR}}{w_C}, \tag{8}$$

where w_{TR} is a satellite transponder bandwidth, w_C - individual channel bandwidth (including guard bands).

The lower value of n_p or n_W determines the capacity ***C*** of the system.

The representative FDMA wireless cellular standards include Advanced Mobile Phone System (AMPS) in the United States, Nordic Mobile Telephones (NMT) in Europe, and Total Access Communications System (TACS) in the United Kingdom [3]. Long Term Evolution [4] (LTE) implements Orthogonal Frequency Division Multiplexing (OFDM) for its downlink and Single-Carrier Frequency-Division Multiple Access (SC-FDMA) for its uplink.

3. Time Division Multiple Access

Time Division Multiple Access (TDMA) is another multiple access technique employed in the digital wireless communication systems. This method has been applied in the late 1980s. The frequency band for the each station is the same, and the separation of the carriers in the transponder is done in time (time intervals), and only

[3] Kiseon Kim, Insoo Koo, CDMA Systems Capacity Engineering. Artech House INC., 2005, 201 pp.

[4] 3rd Generation Partnership Project Long Term Evolution – standard improving Universal Mobile Telecommunications System (UMTS)

one user is allowed to either transmit or receive the information data in each interval (slot).

Figure 3 shows a principle of the TDMA process as three ground stations accessing a single frequency translation satellite transponder. Defined time frames are formatting all the transmission time. Each frame is divided into time intervals (time slots), which are at least as much as possible access. Equal time intervals in each frame are designed for individual stations. Each station can transmit in a specific time interval (slot) for its uplink transmission of a burst of data in sequence and each station has use of the full transponder bandwidth during its time slot. The time slot can be changed on demand. Guard bands are used between the time slots to avoid interference.

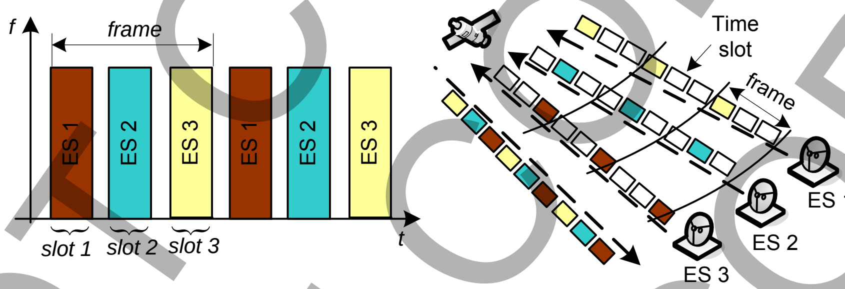


Figure 3. Principle of the Time Division Multiple Access

The information that has been sent in a time slots is called sequence. During an active connection the information is transmitted as sequence, not continuously. At each time frame can be transmitted synchronization, control and user information. TDMA is a convenient method for transmitting digital data because by using digital buffer memory it is converted into time slots and then the receiver restores them. The required capacity of the buffer is determined by:

$$c_{Buf} = r_b . t_F \tag{9}$$

where r_b is a bite rate, t_F is a period of the time slots.

Downlink transmission consists of interleaved set of packets from all the ground stations. A receiving station detects and demultiplexes bursts and delivers information to the users. A reference station is used to establish the synchronization reference clock and provide burst time operational data to the network.

Figure 4 shows a typical time frame. The range of the frame is from 1 to 20 ms. Each station burst contains a *preamble* and *traffic data*. The preamble contains synchronization and station identification data. The *reference burst*, from the reference station, is usually at the start of each frame, and provides the network synchronization and operational information.

Station bursts do not need to be identical in duration. Its can be longer for heavier traffic stations or during higher use periods. The specific allocation of burst times for each of the stations within the frame is called the *burst time plan*. The burst time plan can be changed to adapt for changing traffic patterns.

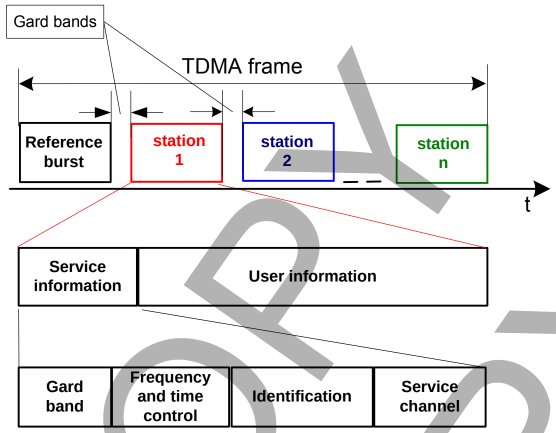


Figure 4. Example of the typical time frame.

Advantages of TDMA:

- TDMA can easily adapt to transmission of data as well as voice communication; it has an ability to carry 64 kbps to 120 Mbps of data rates and this allows the operator to do services like multimedia and videoconferencing, fax, voice;
- TDMA provides no interference from simultaneous transmissions because it separates users according to time;
- TDMA provides users with an extended battery life, since it transmits only portion of the time during conversations;
- less stringent power control due to reduced interference between users;
- TDMA is the most cost effective technology to convert an analog system to digital.

Disadvantages of TDMA:

- higher synchronization overhead;
- the user might be disconnected - because of the users has a predefined time slot when they moving from one cell site to other, if all the time slots in this cell are full;
- the user could not receive a dial tone, if all the time slots in the cell in which the user is currently in are already occupied;
- multipath distortion - to overcome distortion, a time limit can be used on the system, once the time limit is expired the signal can be ignored.

3.1. Time Division Multiple Access with PCM/TDM/PSK application.

One of the most common structures found applications in VSAT networks, consists of a PCM or ADPCM baseband formatting technique (period of the frame has to be multiple of $125\mu s$), TDM multiplexing and modulation by binary PSK or quadrature QPSK. The information is uploaded to one carrier, covering the entire transponder bandwidth. The data rates are usually $60 Mbps$ when the bandwidth is $36 MHz$ or $130 Mbps$ when the bandwidth is $72 MHz$. The synchronization of the frame is maintained by using the codeword includes a sequence of 24 to 48 bits. The codeword is repeated in each frame.

The preamble consists of some components, each with a specific purpose in the TDMA process. Typical components of the TDMA preamble and reference burst are summarized in Table 1 for a representative operational system, the INTELSAT TDMA system, deployed on many early INTELSAT satellites[5]. The INTELSAT TDMA system consists of two reference bursts per frame and operates with a 2 ms frame period.

Table 1. Intelsat TDMA preamble and reference burst structure.

Components		Description	Number of bits
		Preamble	
CBR	Carrier and bit-timing recovery	synchronizing signal for detector	352
UW	Unique word	burst code word	48
TTY	Teletype	operational data communications between stations	16
SC	Service channel	carries network protocol and alarm messages	16
VOW (*2)	Voice order wire	voice communications between stations	2*64
		Reference Burst	
CBR	Carrier and bit-timing recovery	synchronizing signal for detector	352
UW	Unique word	burst code word	48
TTY	Teletype	operational data communications between stations	16
SC	Service channel	carries network protocol and alarm messages	16
VOW (*2)	Voice order wire	voice communications between stations	2*64
CDC	Coordination and delay channel	used to transfer acquisition, synchronization, control, and monitoring info to stations	16

Total number of bits in the preamble is 560, the reference burst consists 576 bits.

3.2. TDMA Frame Efficiency

The efficiency of the TDMA frame is defined as the ratio of number of bits available for traffic and total number of bits in frame[6]

[5] Dennis Roddy, Satellite communications, McGraw-Hill Professional, 2001, 569 pp.

[6] Ippolito, Louis J., Satellite communications systems engineering: atmospheric effects, satellite link design, and system Performance, JohnWiley & Sons Ltd, 2008, 396 pp.

$$\eta_F = \frac{\textit{number of bits available for the traffic}}{\textit{total number of bits in the frame}} = 1 - \frac{\textit{number of overhead bits}}{\textit{total number of bits in the frame}} \qquad (10)$$

$$\eta_F = 1 - \frac{n_r.b_r + n_t.b_p + (n_r + n_t)b_g}{r_T.t_F}, \qquad (11)$$

where t_F is the TDMA frame time, in *s*; r_T is the total TDMA bit rate, in *bps*; n_r - number of reference stations; n_t - number of traffic bursts; b_r - number of *bits* in reference burst; b_p - number of *bits* in the preamble; b_g - number of *bits* in guard band.

The frame efficiency shows how mach of the frame is in used. It will change when increase the total number of bits (a longer frame time) or lowering the overhead in the frame (no traffic bits). The optimum operating structure occurs by providing the longest possible frame time with the lowest total number bits allocated to overhead functions, but the longest frame time needs of largest memory buffers, which increases the overall delay of the signal.

Example 1. How calculate the frame efficiency, if TDMA frame has a frame shows in Table 1 with length $2ms$, a total number of bits in the preamble is $b_p = 560 bits$, in the reference burst - $b_r = 576 bits$, number of bits in guard band - $b_g = 206 bits$, there are two reference stations, each transmitting a reference burst in the frame $n_r = 2$.

Evaluate the TDMA network in terms of the maximum number of traffic terminals and the operating TDMA date rate for a desired minimum frame efficiency of $\eta_F = 0{,}95$.

From Equation 11 the frame efficiency is

$$\eta_F = 1 - \frac{n_r.b_r + n_t.b_p + (n_r + n_t)b_g}{r_T.t_F} = 1 - \frac{2.576 + n_t.560 + (2 + n_t)206}{r_T.0{,}002} = 1 - \frac{1564 + 766 n_t}{r_T.0{,}002}.$$

1) If the TDMA data rate is set at 120 Mbps, the number of terminals that could be supported at a 95% frame efficiency is

$$0{,}95 = 1 - \frac{1564 + 766 n_t}{120.10^6.0{,}002} \Rightarrow n_t = 13{,}6; \quad n_t \approx 13 \text{ traffic terminals.}$$

2) for a network with a 13 traffic terminals (13 time slots) and $\eta_F = 0{,}95$ it can be achieved data rate of

$$0{,}95 = 1 - \frac{1564 + 766.13}{r_T.0{,}002} \Rightarrow r_T = 115{,}22 Mbps.$$

The result chows, if the number of traffic terminals is fixed, network performance can be optimized by setting the data rate r_T at the minimum value to achieve the desired efficiency η_F. Therefore the network ground terminals can operate with variable TDMA data rates.

3.3. TDMA capacity

To evaluate a capacity for any type of data source bit stream (voice, video, data or combination of these) the network channel capacity for a TDMA network can be evaluated in terms of an *equivalent voice-channel capacity*, n_C. The equivalent voice channel capacity is defined as ratio of available information bit rate, r_i and equivalent voice channel bit rate, r_C

$$n_C = \frac{r_i}{r_C} \tag{12}$$

The available information bit rate, r_i, represents that part of the total bit rate available for information (the total bit rate minus the bit rate allocated to overhead functions). The equivalent voice channel bit rate is usually defined as the standard PCM bit rate $r_C = 64kbps$.

Order for determination of TDMA capacity:

1) Determine the composite carrier-to-noise ratio available on the RF link $(C/N_0)_T$.

2) Determine the carrier-to-noise ratio required to achieve the threshold BER desired for the TDMA network

$$\left(\frac{C}{N}\right)_t = \left(\frac{E_b}{N_0}\right)_t + R_T - W_N + M_i + M_A,\ \ dB, \tag{13}$$

where R_T is the TDMA data rate at the desired frame efficiency, η_F, in dB; N_0 - power flux density for a unit bandwidth, in dB; W_N - noise bandwidth of carrier, in dB; C - power of carrier, in dB; $\left(\frac{E_b}{N_0}\right)_t$ - the ratio, required for the threshold BER; M_i - modem implementation margin (the deviation of modem performance from the ideal), in dB, $\approx (1 \div 3)dB$; M_A - adjacent channel interference margin $\approx (1 \div 2)dB$.

3) Determine the TDMA data rate from

$$\left(\frac{C}{N}\right)_t \geq \left(\frac{C}{N}\right)_T. \tag{14}$$

4) Determine the TDMA capacity from (12)

The frame parameters for the reference bursts, traffic bursts, and guard bands defined are: t_F - the TDMA frame time, in s; r_T - the total TDMA bit rate, in *bps*; n_r - number of reference stations; n_t - number of traffic bursts; b_r - number of *bits* in reference burst; b_T - number of *bits* in total TDMA frame; b_p - number of *bits* in the preamble; b_g - number of *bits* in guard band.

Therefore, it can determinate the follow bit rates:

Total TDMA Bit Rate : $r_T = \frac{b_T}{t_F}$

Reference Burst Bit Rate : $r_p = \frac{b_p}{t_F}$

Reference Burst Bit Rate : $r_r = \frac{b_r}{t_F}$

Guard Time Bit Rate : $r_g = \dfrac{b_g}{t_F}$

The available bit traffic rate is

$$r_i = r_T - n_r(r_r + r_g) - n_t(r_p + r_g) \qquad (15)$$

Therefore

$$n_C = \frac{r_i}{r_C} = \frac{r_T - n_r(r_r + r_g) - n_t(r_p + r_g)}{r_C} \qquad (16)$$

This result provides the number of equivalent voice channels that can be supported by the TDMA network for the specified TDMA bit rate, TDMA frame efficiency, and frame parameters.

Example 2. To calculate the TDMA capacity in number of equivalent voice channel n_C with data from Example 1. There is standard PCM bit rate $r_C = 64kbps$ and first level of digital plesiochronous hierarchy structure $r_C = 2048kbps$.

First define the following data rates in *bps*:

Total TDMA Bit Rate : $r_T = \dfrac{b_T}{t_F} = 115{,}22Mbps$;

Reference Burst Bit Rate : $r_p = {b_p}/{t_F} = \dfrac{560}{0{,}002} = 280kbps$;

Reference Burst Bit Rate : $r_r = {b_r}/{t_F} = \dfrac{576}{0{,}002} = 288kbps$;

Guard Time Bit Rate : $r_g = {b_g}/{t_F} = \dfrac{206}{0{,}002} = 103kbps$.

The available bit traffic rate

$$r_i = r_T - n_r(r_r + r_g) - n_t(r_p + r_g) = 115{,}22.10^6 - 2.(288+103).10^3 - 13.(280+103).10^3 = 112{,}818Mbps$$

For standard PCM bit rate $r_C = 64kbps$ the number of equivalent voice channel is:

$$n_C = \frac{r_i}{r_C} = \frac{112{,}818.10^6}{64.10^3} = 1762{,}8; \Rightarrow n_C = 1762\,channels.$$

For bit rate $r_C = 2048kbps$ the number of equivalent voice channel is:

$$n_C = \frac{r_i}{r_C} = \frac{112{,}818.10^6}{2048.10^3} = 55{,}1; \Rightarrow n_C = 55\,channels.$$

This number of equivalent voice channels can be supported to maintain a frame efficiency of $\eta_F = 0{,}95$.

3.4. Switching in satellite TDMA.

Switching in satellite TDMA (Satellite Switched TDMA) lies in the rapid reconfiguration of the diagram of satellite antenna to improve the possibilities for providing additional access to traditional TDMA system. Satellite Switched TDMA adds antenna beam switching to provide additional multiple access capability to adapt to changing demand requirements. The on-board switching is accomplished at intermediate frequency with an $n \times n$ switch matrix. Switching is done in synchronization with the TDMA bursts from the ground stations.

Figure 5 shows example for configuration of a 3×3 satellite switched TDMA architecture. The network consists of three regional beams, designated as A, B, and C beams. The switch matrix mode is shown on the right of the figure, labeled Mode 1, Mode 2, or Mode 3. The options for switching in this case are $n_S = 6$ (showing only three of them). Table 2 shows the number of full versions of switching time for this example.

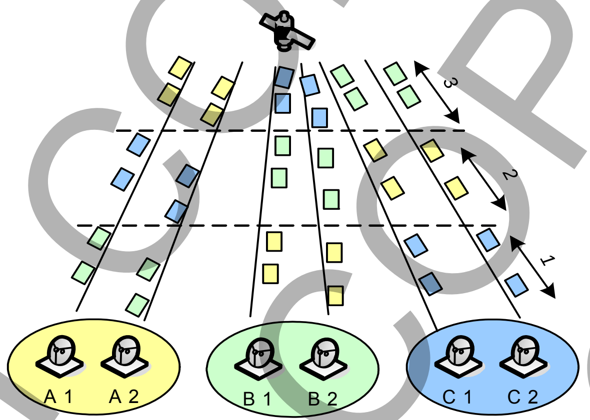


Figure 5. Example of a 3×3 satellite switched TDMA architecture

Table 2. The number of full versions of switching time for 3×3 satellite switched TDMA architecture.

Uplink beam	**Downlink beam**					
	Switch position 1	Switch position 2	Switch position 3	Switch position 4	Switch position 5	Switch position 6
A	B	C	A	A	B	C
B	A	B	C	B	C	A
C	C	A	B	C	A	B

For N number of Regions (beams) $n_S = N!$. As N increases, the number of switch positions becomes quite large, for example:

$N = 3$: $n_S = 6$;

$N = 4$: $n_S = 24$;

$N = 5$: $n_S = 120 \ldots$

The switch consists of coaxial cables or waveguides. Switching elements are ferrites, diodes or FET transistors and dual-gate FET. The architecture is usually limited to three or four beams. For larger matrices grow much weight and dimensions of the switching matrix.

The major TDMA standards contain Global System Mobile (GSM) in Europe and Interim Standard 54/136 (IS-54/136) in North America[7]. GSM was developed in 1990 for second generation (2G) digital cellular mobile communications in Europe. Systems based on this standard were first deployed in 18 European countries in 1991. By the end of 1993, it was adopted in nine more European countries, as well as Australia, Hong Kong, much of Asia, South America, and now the United States.

4. Code Division Multiple Access

The third technique Code Division Multiple Access (CDMA) uses basic principles of the two previous methods. It is a combination of both frequency and time separation. The station has all the bandwidth as in TDMA all in the time, as in FDMA.

CDMA, and. The main characteristics of the CDMA are:

- Works with digital formatted data only;
- Offers the highest power and spectral efficiency operation of the three fundamental techniques;
- All users use same frequency and may transmit simultaneously;
- Narrowband message signal multiplied by wideband spreading signal, or codeword;
- Each user has its own pseudo-codeword (orthogonal to others) and receivers detect only the desired codeword, all others appear as noise.

In this method, each signal is transmitted to the satellite using the same carrier frequency and occupies the entire bandwidth of the transponder, but carriers differ in the way of coding, i.e. they are modulated with a separate code. The receiver on the other side must recognize this particular code; otherwise it can not choose the desired signal. Code modulation used in CDMA is different from channel coding and modulation.

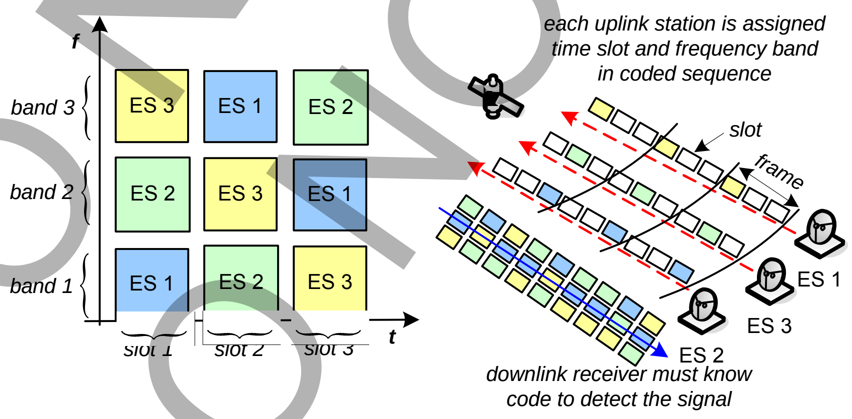


Figure 6. Principle of the Code Division Multiple Access

Figure 6 shows the principle of the Code Division Multiple Access. At each station is given time interval (time slot) and bandwidth. The information is sent to the

[7] Rappaport, T. S., *Wireless Communications*, Englewood Cliffs, NJ: Prentice-Hall, 2002.

satellite ("uplink") as a coded sequence as a series. From satellite to terminals ("downlink") information packages are interleaved, how it is shown in Figure 6. The ground station must know the number and frequency of its time interval to adopt its number of information packets. The ground station separates useful signal from the noise.

The used code sequence has to meet two conditions: to prevent unauthorized access and to be short enough because there can be problems with synchronization or delay of the signal. The most appropriate code sequence that satisfies these conditions is a pseudorandom sequence of binary symbols. It is a binary sequence with a final length in which the position of each bit is random. The autocorrelation of the pseudo-random sequence is similar to the autocorrelation of white noise. Therefore, this sequence is called pseudonoise, PN. What is the difference between random and pseudorandom signal? A random signal cannot be predicted, its future value can be described statistical. A pseudorandom signal is not random at all; it is defined as deterministic periodic signal that is known to both transmitter and receiver. It must comply with three conditions:

- Balance property – in each period of the sequence the number of binary ones differs from the number of binary zeroes by at most one digit;
- Run property - the pseudorandom sequence consist sequences of a same type of binary digits (for example, only ones or only zeroes); the appearance of the opposite digit in a sequence starts a new run; the length of the run is equal of the number of digits in the run;
- correlation property – if the period of the sequence is compared term by term with any cyclic shift of itself, the number of agreements has to differ from the number of disagreements by not more than one count.

The PN sequence used in CDMA systems is generated using sequential logic circuits and a n-stage feedback shift register. Figure 7 shows an example of a 4-stage feedback shift register and table of contents in order of a clock, used to generate the PN sequence. At each clock pulse the binary sequences are shifted through the shift registers one stage to the right. The feedback logic consists of exclusive-OR gates generated by a unique algorithm. The output of the stages are logically combined and fed back as input, generating a PN sequence at the final output. Since the last state 1000 corresponds to the initial one, it is obviously that the register repeats foregoing sequence after 15 clock times. The output sequence of the example on Figure 7 is 000111101011001, where the left digits are the earliest.

The output sequences can be classified as either maximal length (ML) or non maximal length (NML). The number of non-zero states that are possible for this linear PN sequence generator, called its maximal length will be

$$p = ML = 2^n - 1. \tag{17}$$

The sequence has a maximal length if for an n-stage linear feedback shift register the sequence repetition period $\boldsymbol{p}$ is (17).

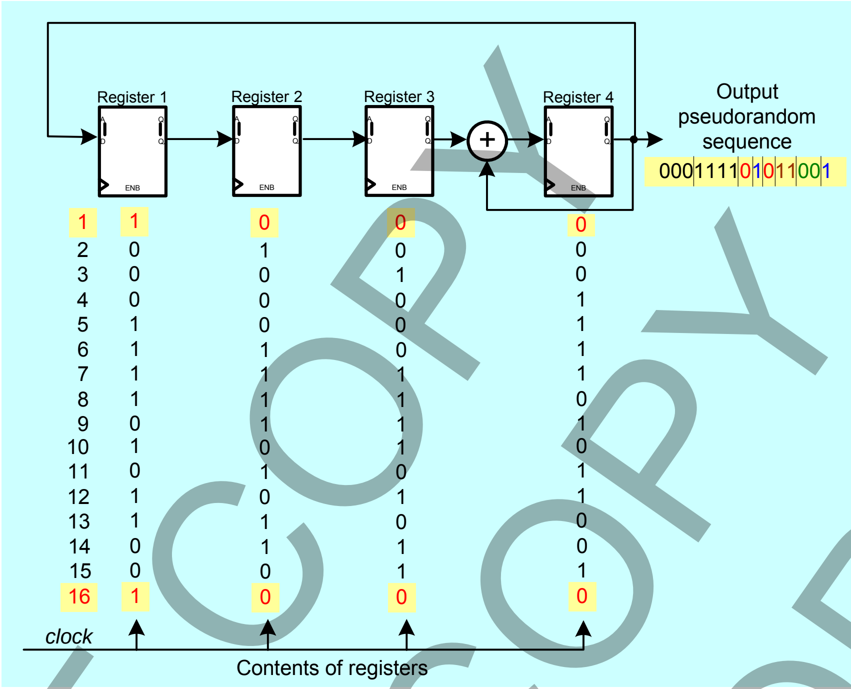


Figure 7. An example of a 4-stage feedback shift register and table of contents in order of a clock

It can test the sequence for the randomness properties. First is balance property. The sequence has seven zeroes and eighth ones, consequently the sequence meets the balance condition. Second condition, the sequence consists of regularly repeated same number (in Fig. 7 they are separated by vertical lines). The number of digits is eight. Four of them (50%) have one digit, two (25%) - two digits, and the other with more than two digits. So the run property is met. Third condition: It was found that normalized autocorrelation function of periodic pulse signal with a period $ML = 2^n - 1$ is as follows:

$$R(\tau) = \frac{1}{ML}.\Delta k\,, \tag{18}$$

where Δk is a difference between the number of agreements and the number of disagreements when comparing part of the sequence and its shifted copy.

It is obviously, when $\tau = 0$ (the signal and its copy match) autocorrelation function is one $R(\tau) = 1$. After shifting the copy off less than ML autocorrelation function is $R(\tau) = -\frac{1}{ML}$. A comparison of the signal and its copy shifted by one register to the right gives the following result:

0	0	0	1	1	1	1	0	1	0	1	1	0	0	1
1	0	0	0	1	1	1	1	0	1	0	1	1	0	0
no	yes	yes	no	yes	yes	yes	no	no	no	no	yes	no	yes	no

There are seven agreements and eight disagreements. From Equation (18)

$$R(\tau) = \frac{1}{15}.(7-8) = -\frac{1}{15}$$

So, the third condition is done. If the copy of the signal is shifted by two or more register (but not more than 15) is obtained the same result. The synthesis of an appropriate structure of pseudorandom sequence generator depends on the meeting the conditions for randomness. Not every structure can maximize the length and satisfy these randomness properties.

The generated PN sequence of binary symbols is combined with digital's data to obtain the number of pseudorandom sequence of binary data. Figure 8 shows the process used to generate the PN data stream. Two digital signals are entering in the modulator: the input digital data stream $m(t)$ at a rate $r_b = 1/t_b$ and pseudorandom sequence of binary symbols $p_{PN}(t)$ at a chip rate $r_{CH} = 1/t_{CH}$. A PN cycle is called chip clock.

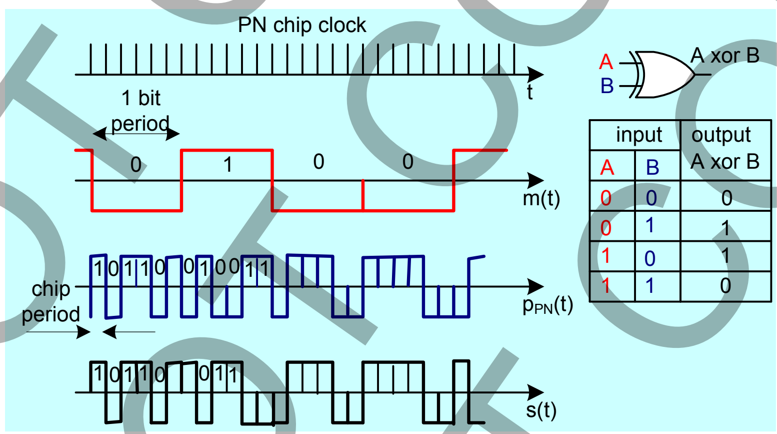


Figure 8. An example of the generation a PN sequence of data

The PN sequence is modulo-2 added to the data sequence m(t) to produce the data stream

$$s(t) = m(t) \oplus p_{PN}(t)\,. \tag{19}$$

Figure 8 shows that the PN data stream is at the chip rate $r_{CH} = 1/t_{CH}$, which is higher than the original data rate $r_b = 1/t_b$, that is $r_{CH} >> r_b$. In the example shown on the figure $r_{CH} = 6r_b$.

The condition $r_{CH} >> r_b$ is very important for the successful implementation of CDMA. It is the reason that CDMA is often referred to as spread spectrum or spread spectrum multiple access, because the original data sequence is “spread” out over a much greater frequency band in the transmission channel.

The chip rate is selected so as to extend the signal over the total available channel bandwidth. Large spreading ratios are typical, for example, in mobile satellite voice networks, the original 16 kbps voice data stream may be spread at a chip rate to produce a PN data stream that operates over an 8MHz RF channel bandwidth. This is a spreading factor of 500, assuming 1 bit/Hz modulation such as BPSK.

What is Spreading Factor, SF? The ratio of both chip and bit rates indicates how many times spectrum of producing data stream has extended compared with the input digital data stream and it is called Spreading Factor or a processing gain when is calculated in dB:

$$\mathrm{SF} = \frac{r_{CH}}{r_b}; \qquad \mathrm{gain}_{\mathrm{BPSK}} = 10\log\left(\frac{r_{CH}}{r_b}\right), \tag{20}$$

because for a $1bps/Hz$ modulation system, such as BPSK, $W_{RF} \approx r_{ch}$; chip rate $r_{CH} = 1/t_{CH}$, data rate $r_b = 1/t_b$, $t_b = n.t_{ch}$.

For example shown on the figure 8 $r_{CH} = 6r_b$,

$$\mathrm{SF} = \frac{r_{CH}}{r_b} = \frac{6r_b}{r_b} = 6,$$

$$\mathrm{gain}_{\mathrm{BPSK}} = 10\log\left(\frac{r_{CH}}{r_b}\right) = 10\log\left(\frac{6r_b}{r_b}\right) = 10\log 6 = 7{,}78\,dB.$$

For example, the standard IS-95, which was introduced as a second generation (2G) wireless cellular communications in 1990 in North America, and later in Japan and Korea, has a chip rate $r_{ch} = 1{,}2288$ and spreading factor $\mathrm{SF} = 64$.

From equation (19) follow spreading the bandwidth needed for transmission of information. Therefore, this access method is often called spread spectrum multiple access. The spread spectrum is defined as a communication method in which the signal occupies a bandwidth, greater than the minimum required to transmit information and that spreading code is independent of the transmitted information.

There are three basic types of CDMA:

- Direct Sequence Spread Spectrum, DS-SS - digital data directly recode with pseudo-random sequence of binary symbols (Fig. 8);

- Frequency Hopping Spread Spectrum, FH-SS - change of carrier frequency, i.e. useful signals are transmitted on different carrier frequencies generated pseudo manner;

- Time Hopping, TH - intermittent switching time interval defined by pseudo-random sequence.

In commercial communications are primarily used the first two methods. The third method, TH-CDMA is used in cases of intentional spread of disturbance as possible to hide the details of the signal from the enemy.

4.1. Direct Sequence Spread Spectrum

The structure of the system with direct sequence spread spectrum (DS-SS) is shown in Figure 9. It includes a phase modulator, PN code modulator, balanced demodulator with bandpass filter in the receiver and demodulator phase.

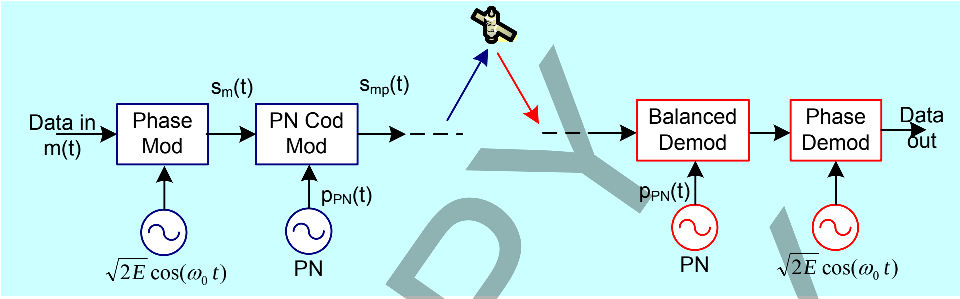


Figure 9. A structure of the system with direct sequence spread spectrum (DS-SS)

The information bitstream is phase modulated onto a carrier (phase modulator), then enter in the PN Code Modulator which phase modulates the RF carrier to produce the spread signal. After passing through the satellite channel, the signal is 'despread' with a balanced modulator, then phase demodulated to produce the original data bitstream.

After phase modulation with a carrier in Phase Modulator input data stream $m(t)$ with the rate $r_b = 1/t_b$ and bandwidth W_{Rb} is as follows:

$$s_m(t) = \sqrt{2E}\cos[\omega_0 t + \psi_m(t)], \quad 0 \le t \le T,$$

where $\psi(t)$ is the information bearing phase modulation.

The PN Code Modulator phase modulates the data modulated signal $s_m(t)$ with the PN sequence $p_{PN}(t)$ with chip rate r_{CH}. The output of the PN modulator is

$$s_{mp}(t) = \sqrt{2E}\cos[\omega_0 t + \psi_m(t) + \psi_p(t)], \tag{21}$$

Thus, the signal to be transmitted to the receiver over the satellite is spread in frequency by the PN sequence to a bandwidth of $W_{R_{CH}}$.

It is known that ideal suppressed carrier binary phase shift keying (BPSK) phase is taking $M = 2$ discrete values, i.e. it switches to 180^0 according to a logic "0" and "1". Then the expression (21) can be written as a multiplication of the carrier and $m(t)$ information antipodes impulses with a values $+1$ or -1:

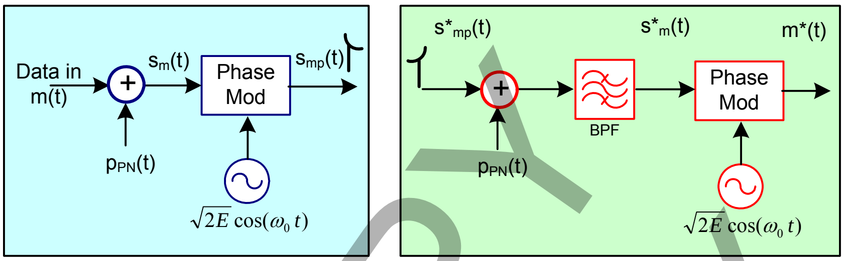
$$s_m(t) = \sqrt{2E}.m(t)\cos(\omega_0 t). \tag{22}$$

The PN sequence $p_{PN}(t)$ is also antipodal pulse stream with a values $+1$ or -1, Equation (22) can be written as

$$s_{mp}(t) = \sqrt{2E}.m(t).p_{PN}(t)\cos(\omega_0 t). \tag{23}$$

The phase of the carrier equals $\psi_m(t) + \psi_p(t) = 180^0$, if the modulo 2 sum of data and code is binary 1 ($m(t) + p_{PN}(t) = 1$); and the phase of the carrier is zero $\psi_m(t) + \psi_p(t) = 0^0$ when the modulo 2 sum of data and code is binary 0 ($m(t) + p_{PN}(t) = 0$).

Figure 10 shows the structure according to Equation (23)



a) b)

Figure 10. An elements of the modulator (a) and demodulator (b) of the direct sequence spread spectrum system

The received spread signal is multiplied by a stored replica of $p_{PN}(t)$ (Fig. 10,b). The output of the balanced demodulator is then

$$s^*_{mp}(t) = \sqrt{2E}.m(t).p^2_{PN}(t)\cos(\omega_0 t)\,. \tag{24}$$

Since the $p_{PN}(t)$ is a binary signal, therefore, $p^2_{PN}(t) = 1$ and

$$s^*_m(t) = \sqrt{2E}.m(t).\cos(\omega_0 t) \tag{25}$$

or

$$s^*_m(t) = \sqrt{2E}\cos[\omega_0 t + \psi^*_m(t)]\,, \tag{26}$$

and the information $\psi^*_m(t)$ can be recovered through the final phase demodulator.

If the receiver PN code differs from that of the transmitter, random phase modulation occurs and spread spectrum signal looks as noise. The bandpass filter (with bandwidth W_{Rb}) removes high frequency sidebands.

Figure 11 shows an example of the direct sequence spread spectrum modulation and demodulation, shown in Figure 10.

In Fig. 11a, b are shown input data sequence $m(t)$ with the bit rate $r_b = 1/t_b$ and bandwidth W_{Rb}, and pseudorandom sequence of binary symbols $p_{PN}(t)$ at a chip rate $r_{CH} = 1/t_{CH}$. In Fig. 11c is shown the result of modulo-2 addition of the data sequence $m(t)$ and code sequence $p_{PN}(t)$. Fig. 11d shows the value of the phase of spread spectrum signal that is π if the modulo 2 sum of data and code is binary 1; and it is a zero when the modulo 2 sum of data and code is binary 0.

The demodulation of a signal begins with dispreading of the spectrum by multiplying the received signal with a synchronized replica of the code $\psi^*_p(t)$ as the antipodal phase shift.

The carrier phase is determined as a result of modulo-2 addition of d) and e) (Fig. 11f). Fig. 11g shows the result of recovering the data waveform by the use of BPSK demodulator.

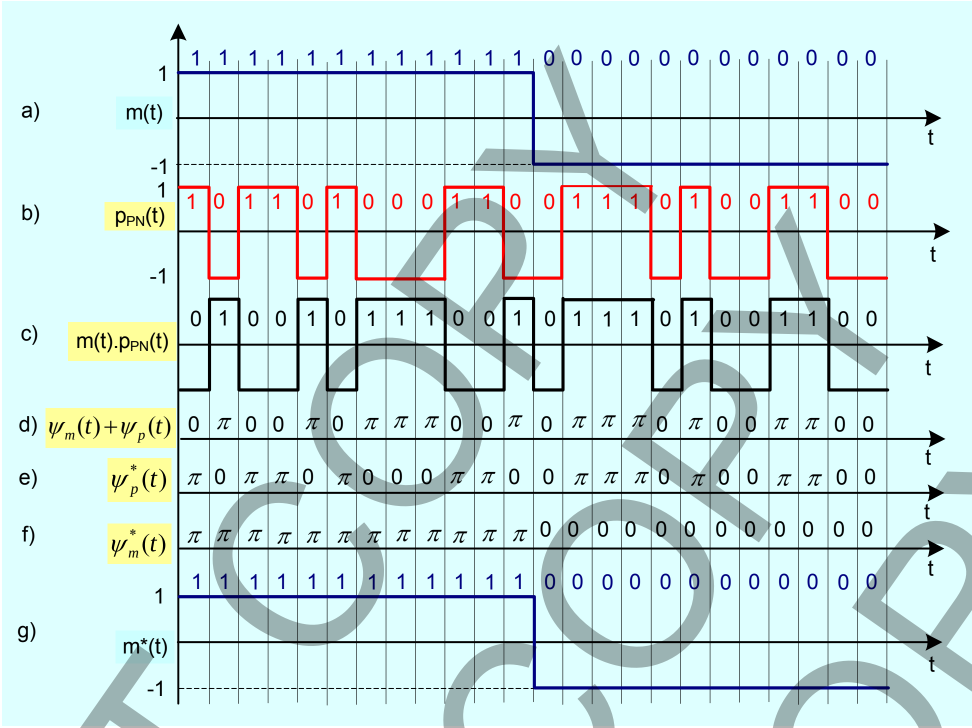


Figure 11. An example of the direct sequence spread spectrum modulation and demodulation

4.2. Frequency Hopping Spread Spectrum.

The main point in Frequency Hopping Spread Spectrum, FH-SS, is that the carrier frequency is changed in accordance of a PN sequence, producing a sequence of modulated data bursts with time varying pseudorandom carrier frequencies.

The all possible carriers frequencies available for frequency hopping in FH-SS is called the *hop set*. Each of the hopped channels contains adequate RF bandwidth for the modulated information, usually a form of frequency shift keying (FSK).

There are two bandwidths defined in FH-SS operation:

- Instantaneous Bandwidth, w_{bb} – the baseband bandwidth of the channel used in the hopset.

- Total Hopping Bandwidth, w_{RF} – the total RF bandwidth over which hopping occurs.

As larger the ratio of w_{RF} to w_{bb}, so better the spread spectrum performance of the FH-SS system. Figure 12 shows the elements of the FH-SS satellite system.

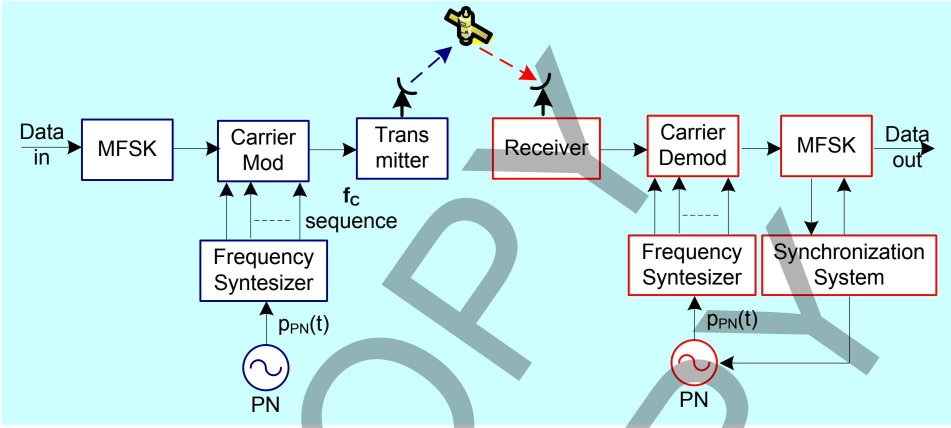


Figure 12. The elements of the FH-SS system.

This technique commonly used M-ary frequency shift keying MFSK where $k = \log_2 M$ information bits are used to determine which frequency of all M is to be transmitted. There are two steps of modulation - data modulation and frequency hopping modulation. The data modulated signal is PN modulated with a PN sequence of carrier frequencies, f_C, generated from the PN sequence $p_{PN}(t)$. The frequency-hopped signal is transmitted through the satellite channel. Receiver reverses the signal processing steps. The received signal is frequency hopping demodulated in the demodulator using a stored replica of the PN sequence. Then the dehopped signal is demodulated by the data demodulator to develop the input data stream. Commonly in used is noncoherent detector because it is difficultly to keep the phase coherence in the frequency hopping.

Example 3. A hopping bandwidth is 840 MHz and frequency step size is $\Delta f = 100 Hz$. What is the minimum number of PN chips required for each frequency word?

The number of frequency are $M = \dfrac{W_{RF}}{\Delta f} = \dfrac{840.10^6}{100} = 84.10^5$;

Minimum number of chips $= \log_2 84.10^5 = 23$.

References

1. Bernard Sklar, Digital Communications: Fundamentals and Applications, Second Edition, Prentice-Hall, 2001.

2. Broadcom Corporation and Cisco Systems, Digital Transmission: Carrier-to-Noise Ratio, Signal-to-Noise Ratio, and Modulation Error Ratio. http://www.broadcom.com/docs/general/Broadcom-Cisco_CNR-SNR-MER.pdf

3. Gilhousen, K. S., et al., On the Capacity of a Cellular CDMA System, *IEEE Trans. On Vehicular Technology*, 1991, pp. 303–312.

4. Dennis Roddy, Satellite communications, McGraw-Hill Professional, 2001, pp.569.

5. Intuitive Guide to Principles of Communications, CDMA: The Concept of signal spreading and its uses in communications, www.complextoreal.com

6. Ippolito, Louis J., Satellite communications systems engineering: atmospheric effects, satellite link design, and system Performance, JohnWiley & Sons Ltd, 2008, pp.396.

7. Kiseon Kim, Insoo Koo, CDMA Systems Capacity Engineering. Artech House INC., 2005, pp.201.

8. Kim, K. I., *Handbook of CDMA System Design, Engineering and Optimization,* Englewood Cliffs, NJ: Prentice Hall, 2000.

9. Koo, I., et al, Sensitivity Analysis for Capacity Increase on the DS-CDMA System, *Proc.of JCCI*, 1997, pp. 447–451.

10.Pickholtz R. L., Schilling D. L., Milstein L. B., Theory of Spread Spectrum Communications: A Tutorial. IEEE Transaction on Communications, Vol. com-30, No. 5, May 1982.

11.Pickholtz R. L., Schilling D. L., Milstein L. B., Revisions to Theory of Spread Spectrum Communications: A Tutorial. IEEE Transaction on Communications, Vol. com-32, No. 2, February 1984.

12.Zhili Sun, Satellite networking principles and protocols, John Wiley & Sons, 2005. pp. 342.

Problems and questions

1. What are the similarities and differences between multiple access and multiplexing?

2. What are the benefits of CDMA versus FDMA and TDMA?

3. Determine the maximum number of users that can access the satellite for the PCM/TDM/PSK/FDMA digital multiple channels per carrier system.

4. Calculate the capacity for PCM/SCPC/PSK/FDMA system.

5. A TDMA system operates on $100Mbps$ with a $2ms$ frame time. Assume that all slots are equal length and guard lines between slots are $1\mu s$. Calculate the efficiency for the case of 1, 5 and 50 slots per frame. Calculate the efficiency again if it is required $500bits$ preamble for each slot.

6. The TDMA frame consists of two reference bursts per frame, with a variable number of traffic bursts, depending on load demand and service area coverage. QPSK modulation (2

bits/symbol) is used, with a total frame length of 120832 symbols. The preamble in each traffic burst is 280 symbols long, the control and delay channel is 8 symbols, and the guard band interval is 103 symbols. Calculate the frame efficiency for a frame consisting of 14 traffic bursts per frame.

7. Determine the voice-channel capacity for the TDMA frame of problem 6.The voice channel is the standard PCM format (64 kbps) with QPSK modulation. The frame period is 2 ms. Assume a speech activity factor of 1.

8. Describe randomness properties that make pseudorandom sequence to be random.

9. What is the average BER for the network with an $\left(\frac{E_b}{N_0}\right) = 3dB$ and $\left(\frac{E_b}{N_0}\right) = 30dB$ if the PN sequence for a BPSK direct sequence spread spectrum network operates at 512 chips per symbol period? The network consists of 150 users with equal received power at the demodulator.

Problems

1. Satellite Communications Segments are:…?
2. Why Geostationary Earth Orbit (GEO) is called Geosynchronous?
3. Find the rang, azimuth angle and elevation angle if the earth station is located at: (a) Coventry, England; (b) Riga, Latvija; (c) Istambul, Turkey. Satellite Longitude is 19.2^0 E.
4. Which type of satellite orbit provides the best performance for each of the following characteristics: (a) Minimum latency (time delay) for voice and data networks; (b) Best coverage of high latitude locations; (c) Minimum free space path loss; (d) Ground terminals with minimum antenna tracking required.
5. The frequencies used for satellite communications are accommodated between…?
6. Why communications satellites also employ polarization frequency reuse?
7. What is a satellite communications link budget?
8. A signal has power 20W, his bit rate is 600bps. What is the bit energy in decibels?
9. Describe the relation between C/N , C/N_0 and E_b/N_0 .
10. The transmitted power of Earth station is $20W$, a distance to satellite is $d = 37500km$. The antenna gain is $45dB$. What is the power flux density?
11. How the values of antenna gain is depending on frequency and antenna's diameter?
12. The frequency uplink is 6,175GHz, a distance to satellite is $d = 37500km$. What is the free space path loss?
13. What is the power flux density and power of received signal if both the transmit and receive parabolic antennas have a diameter of 3 m, the transmit power is 15 watts, the antenna efficiency is 55% for both antennas. The satellite is in a GSO location, with a range of $d = 37500km$. The frequency of operation is 6 GHz.
14. A receiver with a low noise amplifier with a gain of $G_{RF} = 23dB$ and noise temperature of $t_{RF} = 50K$; a downconverter with a gain of $G_M = -10dB$ and noise temperature of $t_M = 500K$, an intermediate frequency unit: $t_{IF} = 1000K$, $G_{IF} = 30dB$ .What is the receiver system noise temperature.
15. Calculate Receiver G/T (dB/K) of a satellite having antenna gain 45 , over all receiver noise temperature $t_S = 70K$.
16. A fixed service satellite uplink operating at 14 GHz consists of a 3m antenna diameter ground terminal with a $1kW$ transmitter. The receiver on the GSO satellite has a 2m antenna and a receive system noise temperature of 500 K. The free space path loss for the link is 202,5 dB. Assume an efficiency of 55%

for both antennas. What is the carrier-to-noise density for the uplink under clear sky conditions with a gaseous attenuation loss of 2,5 dB on the link?

17. There is a satellite with a range $d = 37500km$; transmitted power $p_T = 2W$; antenna gain $g_T = 22\ dB$; channel system noise temperature $t_S = 160K$ and bandwidth $b_N = 500MHz$ of the satellite channel; Frequency = 12GHz. What are: (a) Power Flux Density to the received terminal; (b) Received power to the received terminal with antenna effective aperture $A_{eff} = 10m^2$; (c) Received antenna gain g_R; (d) Received ratio c/n.

18. A satellite network operates with a frequency translation transponder and provides a 64 kbps BFSK data link. The requirement for the link BER is 5.10^{-5}. (a) What is the required composite E_b/N_0 for the link? (b) The downlink C/N_0 for the link is 60 dBHz. What uplink C/N_0 would be required to maintain the BER requirement?

19. What are the main differences, advantages and disadvantages between: (a) FDMA and TDMA; (b) TDMA and CDMA?

20. A communications satellite transponder with a 42MHz usable bandwidth operates with multiple FDMA carriers. Each FDMA carrier requires a bandwidth of 7,5MHz and an EIRP of 16 dBw. The total available EIRP for the link is 24 dBw. Determine the maximum number of carriers that can access the wireless link if assume 10% guard bands and neglect implementation margins.

21. A direct-sequence spread spectrum CDMA satellite network operates with 8 kbps voice channels. The interfering noise density on the network is measured as 6 dB above the thermal noise level. Determine the processing gain for a 6MHz spread spectrum bandwidth.

22. A hopping bandwidth is 820 MHz and frequency step size is $\Delta f = 50Hz$. What is the minimum number of PN chips required for each frequency word?

23. The Hamming distance between sequences $A = 00101101$ and $B = 10110101$ is …?. The Hamming weight of a sequence $A = 00101101$ is…?

24. What is the symbol error correcting capabilities of a RS(7,3) code. Now many bits are there per symbol?

25. What is the codeword polynomial of a message that consist three symbols 010 110 111 and has to be coded by RS(7,3) code with the generator polynomial $g(X) = \alpha^3 + \alpha^1 X + \alpha^0 X^2 + \alpha^3 X^3 + X^4$?

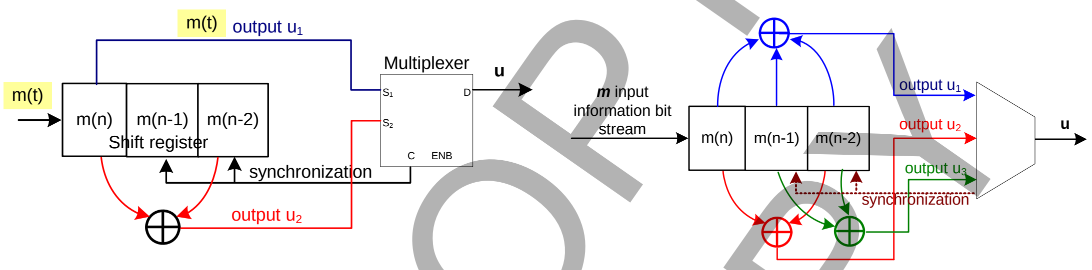
26. Use the generator polynomial for the RS(7,3) code to encode the massage 10110101.

27. Convolutional codes are commonly characterized by parameters…?

28. Draw the trellis diagram for the K=3, code rate $r = 1/3$ code generated by $g_1(x) = 1 + x + x^2$, $g_2(x) = 1 + x^2$ and $g_3(x) = x + x^2$.

29. Describe the generator polynomials of the coder in Figure P1(a) and (b).

30. What is the output coded message generated by encoders, shown in Figure P1 (a) and (b)?



a) b)

Figure P1. Coders of convolutional code

31. What is a trellis diagram for the encoder given in Figure P1(b) and input message $m = 101100$.

32. Choice the path in trellis diagram for the output coded message generated by encoder, shown in Figure P1(b).

33. What are the main differences, advantages and disadvantages between recursive and nonrecursive systematic convolutional encoder.