University of Anbar

College of Engineering

Department of Electrical Engineering

Advanced Communications Systems

EE4335

Dr. Naser Al-Falahy

WEEK 1

SATELLITE COMMUNICATIONS

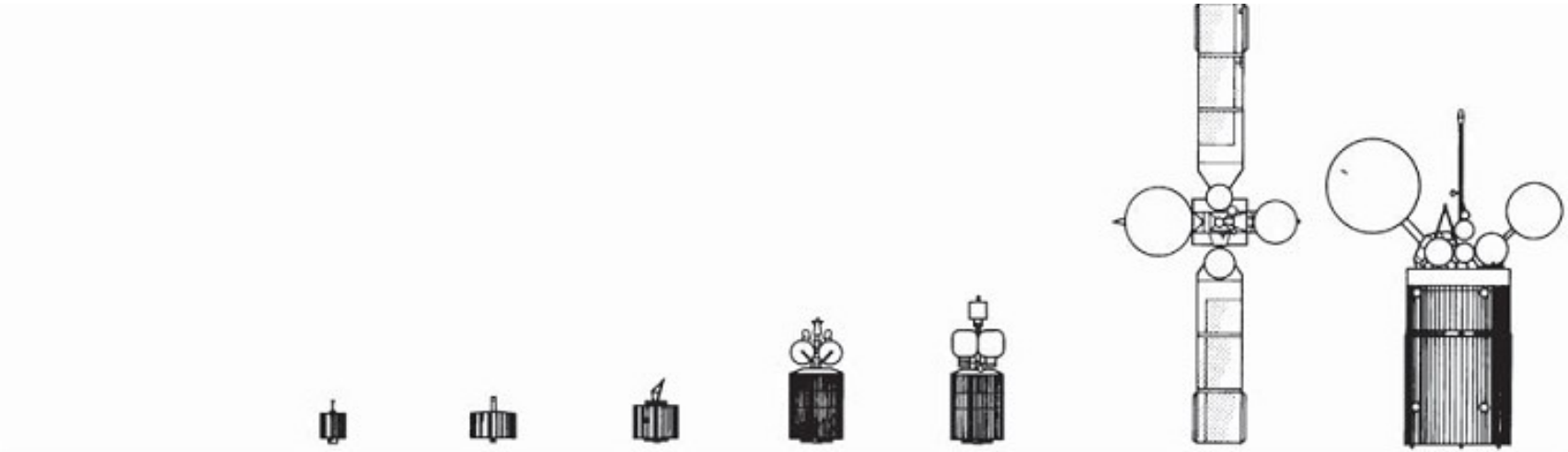
The idea of a communication through a satellite, in particular with a synchronous satellite was conceived by Arthur C. Clarke, a famous British science fiction writer in 1945. Clarke had already pointed out that a satellite in a circular equatorial orbit with a radius of about 42,242 km would have an angular velocity that matched the earth's. Thus, it would always remain above the same spot on the ground and it could receive and relay signals from most of a hemisphere. Three satellite spaced 120 degree apart could cover the whole world with some overlap provided that messages could be relayed between satellites and thus reliable communication between any two points in the world was possible. Clarke had also stated that the electrical power for the satellite would be obtained by conversion of the sun's radiation by means of solar cells. Clarke's paper went almost totally unnoticed until man-made satellites became a reality with Sputnik I (October 4, 1957). However, it may be noted that the synchronous orbit was not achieved until 1963.

Structure of Satellite Communications System

Communications Satellites are usually composed of the following subsystems:

- Communication Payload, normally composed of transponders, antenna, and switching systems.
- Engines used to bring the satellite to its desired orbit
- Station Keeping Tracking and stabilization subsystem used to keep the satellite in the right orbit, with its antennas pointed in the right direction, and its power system pointed towards the sun.
- Power subsystem, used to power the Satellite systems, normally composed of solar cells, and batteries that maintain power during solar eclipse.
- Command and Control subsystem, which maintains communications with ground control stations. The ground control earth stations monitor the satellite performance and control its functionality during various phases of its life-cycle.

The bandwidth available from a satellite depends upon the number of transponders provided by the satellite. Each service (TV, Voice, Internet, radio…etc) requires a different amount of bandwidth for transmission. This is typically known as link budgeting.



Designation: Intelsat	I	II	III	IV	IV A	V	V A/V B	VI
Year of first launch	1965	1966	1968	1971	1975	1980	1984/85	1986/87
Prime contractor	Hughes	Hughes	TRW	Hughes	Hughes	Ford Aerospace	Ford Aerospace	Hughes
Width (m)	0.7	1.4	1.4	2.4	2.4	2.0	2.0	3.6
Height (m)	0.6	0.7	1.0	5.3	6.8	6.4	6.4	6.4
Launch vehicles		Thor Delta		Atlas-Centaur		Atlas-Centaur and Ariane	Atlas-Centaur and Ariane	STS and Ariane
Spacecraft mass in transfer orbit (kg)	68	182	293	1385	1489	1946	2140	12,100/3720
Communications payload mass (kg)	13	36	56	185	190	235	280	800
End-of-life (EOL) power of equinox (W)	40	75	134	480	800	1270	1270	2200
Design lifetime (years)	1.5	3	5	7	7	7	7	10
Capacity (number of voice channels)	480	480	2400	8000	12,000	25,000	30,000	80,000
Bandwidth (MHz)	50	130	300	500	800	2137	2480	3520

Figure 1.1 Evolution of INTELSAT satellites. (*From Colino 1985; courtesy of ITU Telecommunications Journal.*)

INTRODUCTION TO ORBITS

Satellites (spacecraft) orbiting the earth follow the same laws that govern the motion of the planets around the sun. From early times much has been learned about planetary motion through careful observations. Johannes Kepler (1571–1630) was able to derive empirically three laws describing planetary motion. Later, in 1665, Sir Isaac Newton (1642–1727) derived Kepler's laws from his own laws of mechanics and developed the theory of gravitation.

Kepler's laws apply quite generally to any two bodies in space which interact through gravitation. The more massive of the two bodies is referred to as the primary, the other, the secondary or satellite.

Newton's Law of motion can be written as

$$s = ut + \frac{1}{2}\,at^2$$

$$v = u^2 + 2\,at$$

$$F = ma$$

where: s is the distance travelled from t=0, u is the initial velocity at t=0, v is the final velocity at time=t, a is the acceleration, F is the force acting on the body, and m is the mass of the body.

Kepler's First Law

Kepler's first law states that **the path followed by a satellite around the primary will be an ellipse**. An ellipse has two focal points shown as F1 and F2 in Fig.1. The center of mass of the two body system, termed the barycenter, is always centered on one of the foci. In our specific case, because of the enormous difference between the masses of the earth and the satellite, the center of mass coincides with the center of the earth, which is therefore always at one of the foci.

The semi-major axis of the ellipse is denoted by a, and the semi-minor axis, by b. The eccentricity e is given by:

$$e = 2\frac{a-b}{a+b}$$

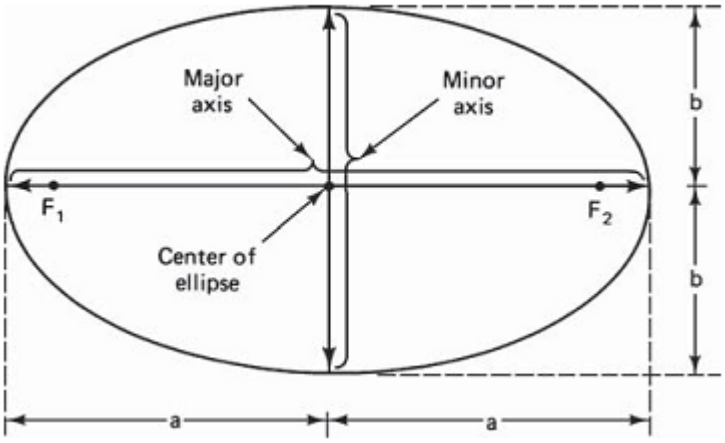


Fig1. The foci F1 and F2 , the semi-major axis a, and the semi-minor axis b of an ellipse.

Note that: e=0 for circular orbits, i.e., a=b

The eccentricity and the semimajor axis are two of the orbital parameters specified for satellites (spacecraft) orbiting the earth. For an elliptical orbit, $0 < e < 1$. When e =0, the orbit becomes circular.

Kepler's Second Law

Kepler's second law states that, **for equal time intervals, a satellite will sweep out equal areas in its orbital plane, focused at the barycenter.**

Referring to fig. 2, assuming the satellite travels distances S1and S2 meters in 1 s, then the areas A1 and A2 will be equal. The average velocity in each case is S1 and S2 m/s, and because of the equal area law, it follows that the velocity at S2 is less than that at S1. An important consequence of this is that the satellite takes longer to travel a given distance when it is farther away from earth. Use is made of this property to increase the length of time a satellite can be seen from particular geographic regions of the earth.

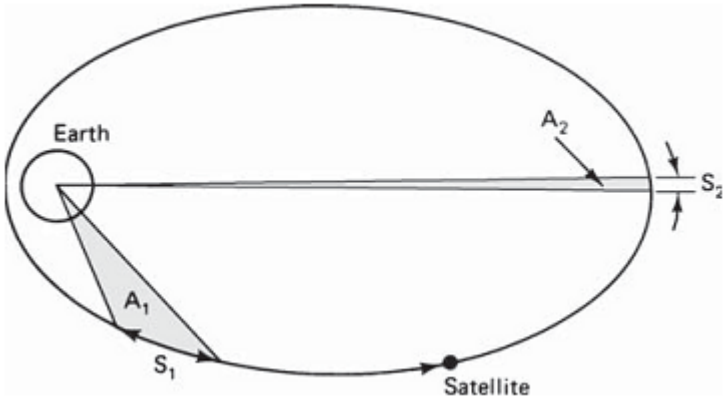


Fig.2 Kepler's second law. The areas A1 and A2 swept out are equal in time.

Kepler's Third Law

Kepler's third law states that **the square of the periodic time of orbit is proportional to the cube of the mean distance between the two bodies.** The mean distance is equal to the semimajor axis a. For the artificial satellites orbiting the earth, Kepler's third law can be written in the form:

$$a^3 = \frac{\mu}{\eta^2}$$

where η is the mean angular velocity of the satellite in radians per second and μ is the earth's geocentric gravitational constant. Its value is:

$$\mu = 3.986005 * 10^{14} \ m^3/s^2$$

The orbital period in seconds is given by:

$$T = \frac{2\pi}{\eta}$$

The importance of Kepler's third law is that it shows there is a fixed relationship between period and semimajor axis. One very important orbit in particular, known as the geostationary orbit, is determined by the rotational period of the earth. In anticipation of this, the approximate radius of the geostationary orbit is determined in the following example.

Example: Calculate the radius of a circular orbit for which the period is 1 day.

Solution: There are 86,400 seconds in 1 day, and therefore the angular velocity/mean motion is:

$$\eta = \frac{2\pi}{86400} = 7.272 * 10^{-5} \ rad/s$$

From Kepler's third law:

$$a = \sqrt[3]{\left[\frac{3.986005 * 10^{14}}{(7.272 * 10^{-5})^2}\right]}$$

$$a = 42.241 \ km$$

Since the orbit is circular, the semi major axis is the same as the radius.

WEEK 2

Earth-Orbiting Satellites

As mentioned previously, Kepler's laws apply in general to satellite motion around a primary body. For the particular case of earth-orbiting satellites, certain terms are used to describe the position of the orbit with respect to the earth.

Sub satellite path. This is the path traced out on the earth's surface directly below the satellite.

Apogee is the point farthest from earth.

Perigee is the point of closest approach to earth.

Apogee and Perigee Heights

Although not specified as orbital elements, the apogee height and perigee height are often required. The length of the radius vectors at apogee and perigee can be obtained from the geometry of the ellipse:

$$r_a = a(1 + e)$$

$$r_p = a(1 - e)$$

Example: Calculate the apogee and perigee heights for the following orbital parameters:

Mean earth radius of R= 6371 km, e =0.0011501 and a =7192.335 km.

Solution:

Using Eqs. above:

r_a =7192.335(1 + 0.0011501) = 7200.607km

r_p =7192.335(1 - 0.0011501) = 7184.063 km

The corresponding heights are:

$$h_a = r_a - R = 829.6\ Km$$

$$h_p = r_p - R = 813.1\ Km$$

FREQUENTLY USED ORBITS

Geostationary Orbits (GEO)

To an observer on the earth, a satellite in a geostationary orbit appears motionless, in a fixed position in the sky. This is because it revolves around the earth at the earth's own angular velocity (360 degrees every 24 hours, in an equatorial orbit).

A geostationary orbit is useful for communications because ground antennas can be aimed at the satellite without their having to track the satellite's motion. This is relatively inexpensive. In applications that require a large number of ground antennas, such as Direct TV distribution, the savings in ground equipment can more than outweigh the cost and complexity of placing a satellite into orbit.

The main drawback of a geostationary orbit is the height of the orbit, usually which requires more powerful transmitters, larger-than-normal (usually dish) antennas, and higher-sensitivity receivers on the earth. The large distance also introduces a significant delay, of ~0.25 seconds, into communications.

For GEO, h≈ 36000 km , i=0^o and e= 0^o.

Since i and e cannot be exactly zero in any practical satellite, the more accurate term geosynchronous orbit is frequently used.

For small inclinations i< 0.5^o, the satellite appears to move in north-south direction slowly in an oscillating fashion with 1-day period.

For small eccentricities; the satellite appears to oscillate in the East-West direction in 1-day period as well. If both oscillations are combined, a **figure of 8 motion** will appear. Station keeping task is to reduce the amplitude of oscillation to small magnitude.

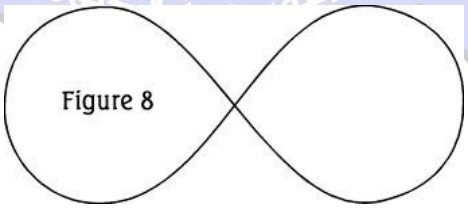


Fig.(3) figure of 8 traced by satellite.

Example

A quasi-GEO satellite is in a circular equatorial orbit close to geosynchronous altitude. The quasi-GEO satellite, however, does not have a period of one sidereal day: its orbital period is exactly 24 h—one solar day. Calculate

(i) the radius of the orbit

(ii) the rate of drift around the equator of the subsatellite point in degrees per (solar) day. An observer on the earth sees that the satellite is drifting across the sky.

(iii) Is the satellite moving toward the east or toward the west?

Sol:

$$T^2 = (4\pi^2 a^3)/\mu$$

Rearranging the equation, the orbital radius a is given by

$$a^3 = T^2\mu/(4\pi^2) = (86{,}400)^2 \times 3.986004418 \times 10^5/4\pi^2$$
$$= 7.5371216 \times 10^{13}\ \text{km}^3$$
$$a = 42{,}241.095\ \text{km}$$

Part (ii) The orbital period of the satellite (one solar day) is longer than a sidereal day by 3 min 55.9 s = 235.9 s. This will cause the subsatellite point to drift at a rate of 360° × 235.9/86400 per day or 0.983° per day.

Part (iii) The earth moves toward the east at a faster rate than the satellite, so the drift will appear to an observer on the earth to be toward the west. ■

Polar Orbiting Satellites

Polar orbiting satellites orbit the earth in such a way as to cover the north and south Polar Regions. (Note that the term polar orbiting does not mean that the satellite orbits around one or the other of the poles).

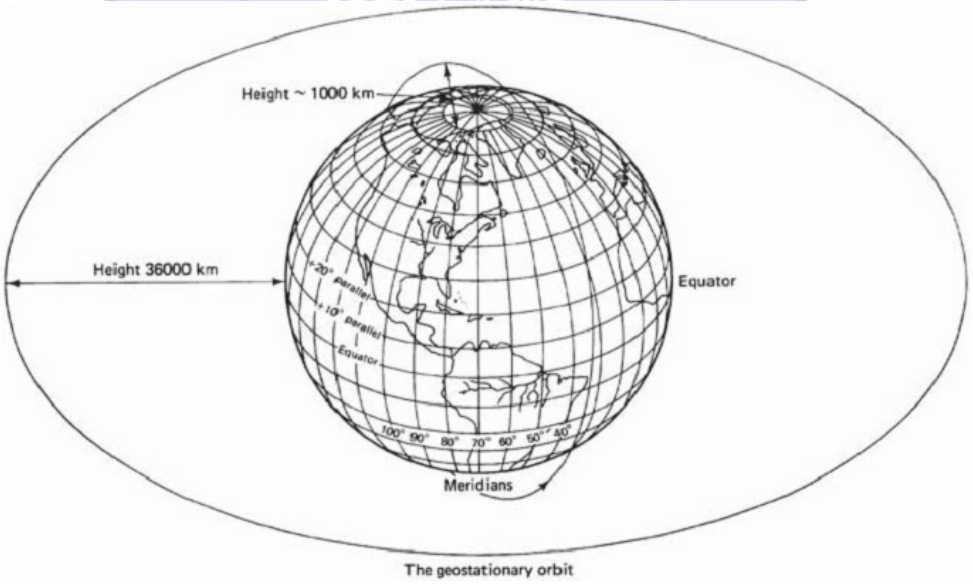


Fig.4 a Geostationary orbit and one possible polar orbit.

Figure 4 shows a polar orbit in relation to the geostationary orbit. Whereas there is only one geostationary orbit, there are, in theory, an infinite number of polar orbits. The U.S. experience with weather satellites has led to the use of relatively low orbits, ranging in altitude between 800 and 900 km, compared with 36,000 km for the geostationary orbit.

Low-Earth-orbit (LEO)

A Low Earth Orbit (LEO) typically is a circular orbit stretch approximately 160 to 1600 km above the earth's surface. In addition, satellites in low earth orbit change their position relative to the ground position quickly. So even for local applications, a large number of satellites are needed if the mission requires uninterrupted connectivity.

Low earth orbiting satellites are less expensive to launch into orbit than geostationary satellites and, due to proximity to the ground, do not require as high signal strength (Recall that signal strength falls off as the square of the distance from the source, so the effect is dramatic). Thus there is a trade-off between the number of satellites and their cost. In addition, there are important differences in the onboard and ground equipment needed to support the two types of missions.

LEOs are subject to aerodynamic drag caused by resistance of the earth's atmosphere to the satellite passage. The exact value of the force caused by the drag depends on atmospheric density, the shape of the satellite, and the satellite's velocity. This force may be expressed in the form:

$$F_d = -0.5\rho_a C_d A_{eq} v^2 \quad kg \cdot m/sec^2$$

where

p_a: atmospheric density. This density is altitude-dependent, and its variation is exponential.
C_d : coefficient of aerodynamic drag.
A_{eq} : equivalent surface area of the satellite that is perpendicular to the velocity,
v : velocity of the satellite.

If the mass ms of the satellite is known, the acceleration a_d due to aerodynamic drag can be expressed as:

$$a_d = \frac{F_d}{m_s} \quad m/sec^2$$

The effect of the drag is a decrease of the orbit's semi-major axis due to the decrease in its energy. A circular orbit remains as such, but its altitude decreases whereas its velocity increases. Due to drag, the apogee in the elliptical orbit becomes lower and, as a consequence, the orbit gradually becomes circular.

The longer the influence on the orbit, the slower the satellite becomes, and it eventually falls from orbit. Aerodynamic drag is more significant at low altitudes (200 to 400 km) and negligible only about 3000 km because, in spite of the low value of atmospheric density encountered at the altitudes of satellites, their high orbital velocity implies that perturbations due to drag are very significant.

- Aerodynamic drag tends to reduce orbital height & eccentricity. It does not affect inclination and it is negligible in geosynchronous satellite.

- The effect is to remove kinetic energy from spacecraft & causing it to fall toward the earth, which in turn increases the orbital velocity resulting in higher drag & faster orbital decay.

Medium and High Earth orbit (MEO) & (HEO)

Medium Earth orbit (MEO), sometimes called intermediate circular orbit (ICO), is the region of space around the Earth above low Earth orbit (altitude of 2,000 kilometers) and below geostationary orbit (altitude of 35,786 kilometers).

The most common use for satellites in this region is for navigation, communication, and geodetic/space environment science. The most common altitude is approximately 20,200 kilometers, which yields an orbital period of 12 hours, as used, for example, by the Global Positioning System (GPS). Other satellites in Medium Earth Orbit include Glonass (with an altitude of 19,100 kilometers) and Galileo (with an altitude of 23,222 kilometers). Communications satellites that cover the North and South Pole are also put in MEO.

The orbital periods of MEO satellites range from about 2 to nearly 24 hours.

- A high Earth orbit (HEO) is a geocentric orbit with an altitude above that of a geosynchronous orbit.
- A highly elliptical orbit (HEO) is an elliptic orbit with a low-altitude (about 1,000 kilometers) perigee and a high-altitude (over 35,786 kilometers) apogee.

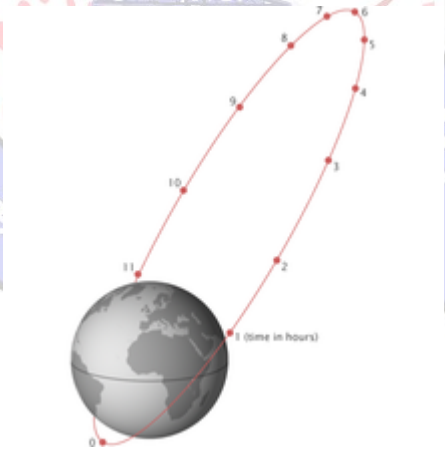
Such extremely elongated orbits have the advantage of long dwell times at a point in the sky during the approach to, and descent from, apogee. Visibility near apogee can exceed twelve hours of dwell at apogee with a much shorter and faster-moving perigee phase. Bodies moving through the long apogee dwell can appear still in the sky to the ground.

Examples of HEO orbits offering visibility over Earth's Polar Regions, Molniya orbits, named after the Molniya Soviet communication satellites which used them.

Molniya Orbit

For areas close to the North (and South) Pole, a geostationary satellite may appear below the horizon. Therefore Molniya orbit satellite has been launched, mainly in Russia, to alleviate this problem. The first satellite of the Molniya series was launched on 1965 and was used for experimental transmission of TV signal from Moscow uplink station to downlink station located in Siberia and the Russian far east.

Molniya (meaning "lightning") was a military communications satellite system used by the Soviet Union. The satellites were placed into highly eccentric elliptical orbits (HEO) known as Molniya orbits, characterized by an inclination of +63.4 degrees and a period of around 12 hours. Such orbits allowed them to remain visible to sites in Polar Regions for extended periods.

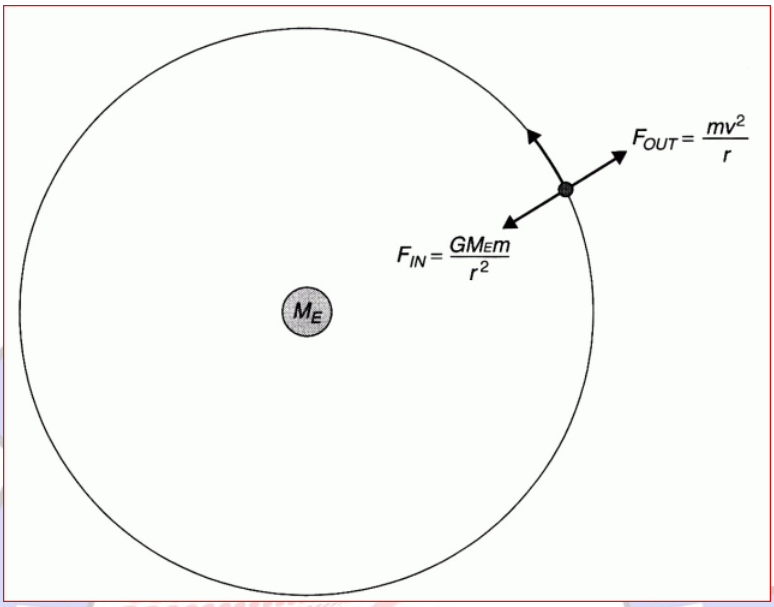


The Molniya orbit is designed so that the satellite spends the great majority of its time over the far northern latitudes. Its period is one half day, so that the satellite is available for operation over the targeted region for six to nine hours every revolution. In this way a constellation of three Molniya satellites (plus in-orbit spares) can provide uninterrupted coverage.

WEEK 3

Satellite Motion

Consider a satellite travelling in a circular orbit, the gravitational force exerted on it by the Earth provides the centripetal force that retains the satellite in its orbit; otherwise the satellite would fly-off in the direction of its tangential velocity.



The centripetal force F is:

$$F = \frac{G\, m\, m_E}{r^2} \ldots\ldots Newton\ Law$$

Also, the centrifugal force is

$$F = \frac{m\, v^2}{r}$$

then

$$\frac{G\, m\, m_E}{r^2} = \frac{m\, v^2}{r}$$

and

$$v^2 = \frac{G\, m_E}{r}$$

Where

G: Gravitational Constant ≈ 6.673×10^{-11} Nm^2/kg^2
m: mass of satellite
R: radius of satellite orbit
m_E: Mass of Earth ≈ 5.98×10^{24} kg.
v: Satellite tangential velocity.

The velocity is independent on satellite mass, but only on the radius of the orbit, using:

$$g_o = \frac{G\, m_E}{R^2}$$

Where g_o is the Earth acceleration = 9.8 m/sec^2
R is the Radius of the Earth=6370 Km.
then:

$$v^2 = \frac{g_o R^2}{r}$$

Example
Calculate the velocity of an artificial satellite orbiting the Earth in a circular orbit at an altitude of 200 km above the Earth's surface. Ans: v = 7,784 m/s

For circular orbit, the distance travelled by a satellite is 2πr. Then, the period of the satellite orbit will be:

$$T=(2\pi r)/v$$

The table below shows the orbital velocity, height and period of four satellite system.

Satellite system	Orbital height (*km*)	Orbital velocity (*km/s*)	Orbital period (*h* *min* *s*)
Intelsat (GEO)	35,786.03	3.0747	23 56 4.1
New-ICO (MEO)	10,255	4.8954	5 55 48.4
Skybridge (LEO)	1,469	7.1272	1 55 17.8
Iridium (LEO)	780	7.4624	1 40 27.0

Mean earth radius is 6378.137 km and GEO radius from the center of the earth is 42,164.17 km.

the orbital speed (v) of a satellite traveling along elliptic orbit can be computed as:

$$v = \sqrt{\mu \left(\frac{2}{r} - \frac{1}{a}\right)}$$

where:

- μ is the standard gravitational parameter.
- r is the distance between the orbiting bodies.
- a is the length of the semi-major axis.

Examples

Example1 GEO

The earth rotates once per sidereal day of 23 h 56 min 4.09 s. Use Eq. (2.21) to show that the radius of the GEO is 42,164.17 km as given in Table 2.1.

Answer Equation (2.21) gives the square of the orbital period in seconds

$$T^2 = (4\pi^2 a^3)/\mu$$

Rearranging the equation, the orbital radius a is given by

$$a^3 = T^2\mu/(4\pi^2)$$

For one sidereal day, $T = 86{,}164.09$ s. Hence

$$a^3 = (86{,}164.1)^2 \times 3.986004418 \times 10^5/(4\pi^2) = 7.496020251 \times 10^{13}\ \text{km}^3$$
$$a = 42{,}164.17\ \text{km}$$

Example2 LEO

The Space Shuttle is an example of a low earth orbit satellite. Sometimes, it orbits at an altitude of 250 km above the earth's surface, where there is still a finite number of molecules from the atmosphere. The mean earth's radius is approximately 6378.14 km. Using these figures, calculate the period of the shuttle orbit when the altitude is 250 km and the orbit is circular. Find also the linear velocity of the shuttle along its orbit.

Answer The radius of the 250-km altitude Space Shuttle orbit is $(r_e + h) = 6378.14 + 250.0 = 6628.14$ km

From Eq. 2.21, the period of the orbit is T where

$$T^2 = (4\pi^2 a^3)/\mu = 4\pi^2 \times (6628.14)^3/3.986004418 \times 10^5\ \text{s}^2$$
$$= 2.88401145 \times 10^7\ \text{s}^2$$

Hence the period of the orbit is

$$T = 5370.30\ \text{s} = 89\ \text{min}\ 30.3\ \text{s}.$$

The circumference of the orbit is $2\pi a = 41{,}645.83$ km.
Hence the velocity of the Shuttle in orbit is

$$2\pi a/T = 41{,}645.83/5370.13 = 7.755\ \text{km/s}$$

Alternatively, you could use Eq. (2.5): $v = (\mu/r)^{1/2}$. The term $\mu = 3.986004418 \times 10^5\ \text{km}^3/\text{s}^2$ and the term $r = (6378.14 + 250.0)$ km, yielding $v = 7.755$ km/s.

Exercises:

1- An artificial satellite is in an elliptical orbit which brings it to an altitude of 250 km at perigee and out to an altitude of 500 km at apogee. Calculate the velocity of the satellite at both perigee and apogee.

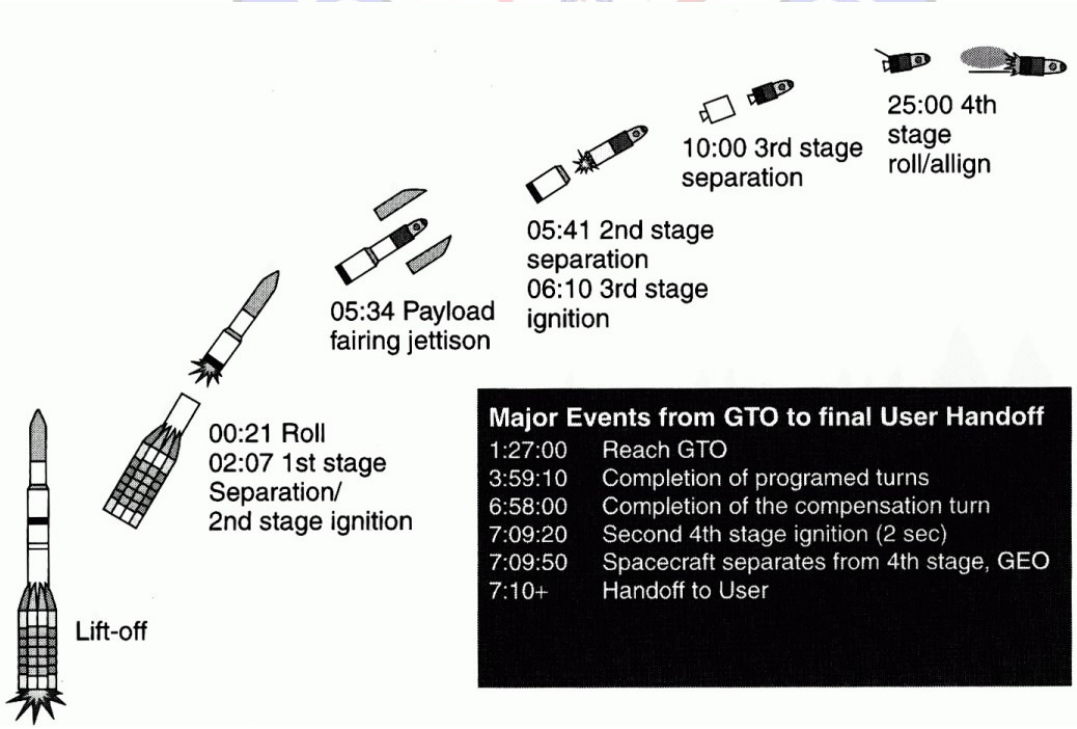
2- Calculate the radius of orbit for a Earth satellite in a geosynchronous orbit, where the Earth's rotational period is 86,164.1 seconds.

3- A satellite in Earth orbit passes through its perigee point at an altitude of 200 km above the Earth's surface and at a velocity of 7,850 m/s. Calculate the apogee altitude of the satellite & the eccentricity.

4- A satellite in Earth orbit has a semi-major axis of 6,700 km and an eccentricity of 0.01. Calculate the satellite's altitude at both perigee and apogee.

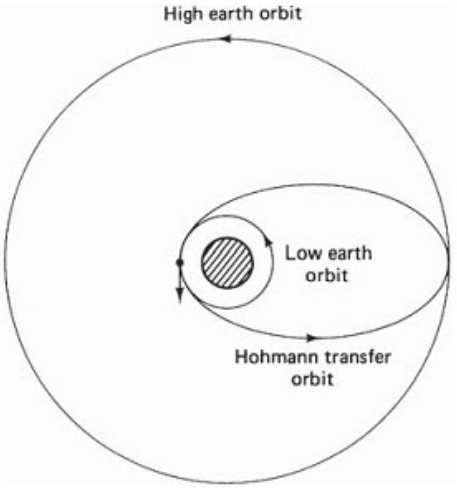
Launching Orbits

Satellites may be directly injected into low-altitude orbits, up to about 200 km altitude, from a launch vehicle. Launch vehicles may be classified as expendable or reusable. Typical of the expendable launchers are the U.S. Atlas-Centaur and Delta rockets and the European Space Agency Ariane rocket. Japan, China, and Russia all have their own expendable launch vehicles, and one may expect to see competition for commercial launches among the countries which have these facilities.



When an orbital altitude greater than about 200 km is required, it is not economical in terms of launch vehicle power to perform direct injection, and the satellite must be placed into transfer orbit between the initial LEO and the final high-altitude orbit. In most cases, the transfer orbit is selected to minimize the energy required for transfer, and such an orbit is known as a Hohmann transfer orbit.

Assume for the moment that all orbits are in the same plane and that transfer is required between two circular orbits, as illustrated in Fig. The Hohmann elliptical orbit is seen to be tangent to the low-altitude orbit at perigee and to the high-altitude orbit at apogee. At the perigee, in the case of rocket launch, the rocket injects the satellite with the required thrust into the transfer orbit. With the STS, the satellite must carry a perigee kick motor (PKM) which imparts the required thrust at perigee. At apogee, the apogee kick motor (AKM) changes the velocity of the satellite to place it into a circular orbit.



It takes 1 to 2 months for the satellite to be fully operational. Throughout the launch and acquisition phases, a network of ground stations, spread across the earth, is required to perform the tracking, telemetry, and command (TT&C) functions.

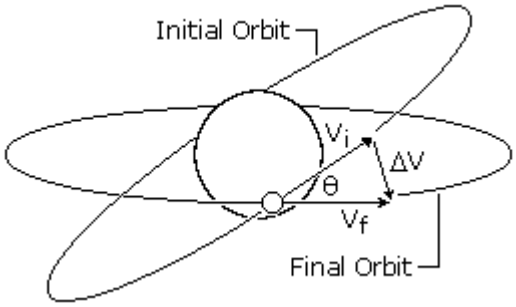
Velocity changes in the same plane change the geometry of the orbit but not its inclination. In order to change the inclination, a velocity change is required normal to the orbital plane. Changes in inclination can be made at either one of the nodes, without affecting the other orbital parameters. Since energy must be expended to make any orbital changes, a geostationary satellite should be launched initially with as low an orbital inclination as possible. It will be shown shortly that the smallest inclination obtainable at initial launch is equal to the latitude of the launch site. Thus the farther away from the equator a launch site is, the less useful it is, since the satellite has to carry extra fuel to make a change in inclination. Russia does not have launch sites south of 45°N, which makes the launching of geostationary satellites a much more expensive operation for Russia than for other countries which have launch sites closer to the equator.

WEEK 4

ORBIT MANOUVERS

- **Orbital plane change (inclination)**

It is an orbital maneuver aimed at changing the inclination of an orbiting body's orbit. This maneuver is also known as an orbital plane change as the plane of the orbit is tipped. This maneuver requires a change in the orbital velocity vector (delta v) at the orbital nodes (i.e. the point where the initial and desired orbits intersect, the line of orbital nodes is defined by the intersection of the two orbital planes).



In general, inclination changes can take a very large amount of delta v to perform, and most mission planners try to avoid them whenever possible to conserve fuel. This is typically achieved by launching a spacecraft directly into the desired inclination, or as close to it as possible so as to minimize any inclination change required over the duration of the spacecraft life.

When both orbits are circular (i.e. e = 0) and have the same radius the Delta-v (Δv_i) required for an inclination change (Δv_i) can be calculated using:

$$\Delta v_i = 2v \sin\left(\frac{\Delta i}{2}\right)$$

where:

- v is the orbital velocity and has the same units as Δv_i
- (Δi) inclination change required.

Example

Calculate the velocity change required to transfer a satellite from a circular 600 km orbit with an inclination of 28 degrees to an orbit of equal size with an inclination of 20 degrees.

SOLUTION,

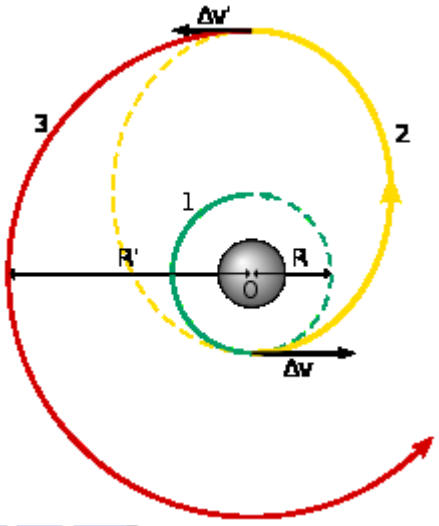
r = (6,378.14 + 600) × 1,000 = 6,978,140 m , ϑ = 28 - 20 = 8 degrees

Vi = SQRT[GM / r]
Vi = SQRT[3.986005×1014 / 6,978,140]
Vi = 7,558 m/s

Δv_i = 2 × Vi × sin(ϑ/2)
Δv_i = 2 × 7,558 × sin(8/2)
Δv_i = 1,054 m/s

- **Orbital altitude change**

In orbital mechanics, the Hohmann transfer orbit is an elliptical orbit used to transfer between two circular orbits of different altitudes, in the same plane.



The orbital maneuver to perform the Hohmann transfer uses two engine impulses, one to move a spacecraft onto the transfer orbit and a second to move off it. This maneuver was named after Walter Hohmann, the German scientist who published a description of it in his book.

The diagram shows a Hohmann transfer orbit to bring a spacecraft from a lower circular orbit into a higher one. It is one half of an elliptic orbit that touches both the lower circular orbit that one wishes to leave (labeled *1* on diagram) and the higher circular orbit that one wishes to reach (*3* on diagram). The transfer (*2* on diagram) is initiated by firing the spacecraft's engine in order to accelerate it so that it will follow the elliptical orbit; this adds energy to the spacecraft's orbit. When the spacecraft has reached its destination orbit, its orbital speed (and hence its orbital energy) must be increased again in order to change the elliptic orbit to the larger circular one.

The total energy of the body is the sum of its kinetic energy and potential energy, and this total energy also equals half the potential at the average distance, (the semi-major axis):

$$E = \tfrac{1}{2}mv^2 + \tfrac{1}{2}Mv_M^2 - \frac{GMm}{r} = \frac{-GMm}{2a}.$$

Solving this equation for velocity results in the vis-viva equation:

$$v^2 = \mu\left(\frac{2}{r} - \frac{1}{a}\right)$$

where:

- v is the speed of an orbiting body
- $\mu = GM$ is the standard gravitational parameter of the primary body, assuming $M+m$ is not significantly bigger than M.
- r is the distance of the orbiting body from the primary focus
- a is the semi-major axis of the body's orbit. = $(r_1 + r_2)/2$

Hence, the energy of the transfer orbit is greater than the energy of the inner orbit (a = r1), and smaller than the energy of the outer orbit (a = r2). The velocities of the transfer orbit at perigee and apogee are given, from the conservation of energy equation, as

$V_{perigee}$ is the perigee velocity & is given by:

$$V_p^2 = 2\mu\left\{\frac{1}{r_1} - \frac{1}{r_1 + r_2}\right\}$$

And the velocity at apogee is given by:

$$V_a^2 = 2\mu\left\{\frac{1}{r_2} - \frac{1}{r_1 + r_2}\right\}$$

As we know, the velocities of the circular orbits are:

$v_1 = \sqrt{\mu/r_1}$ and $v_2 = \sqrt{\mu/r_2}$.

Hence, the required impulses at perigee and apogee are,

$$\Delta v_1 = \sqrt{\frac{\mu}{r_1}}\left(\sqrt{\frac{2r_2}{r_1 + r_2}} - 1\right),$$

$$\Delta v = V_{perigee} - V_{1(circular)}$$

$$\Delta v_2 = \sqrt{\frac{\mu}{r_2}}\left(1 - \sqrt{\frac{2r_1}{r_1 + r_2}}\right),$$

where r_1 and r_2 are, respectively, the radii of the departure and arrival circular orbits.

The total Δv is then:

$\Delta v_{total} = \Delta v_1 + \Delta v_2$.

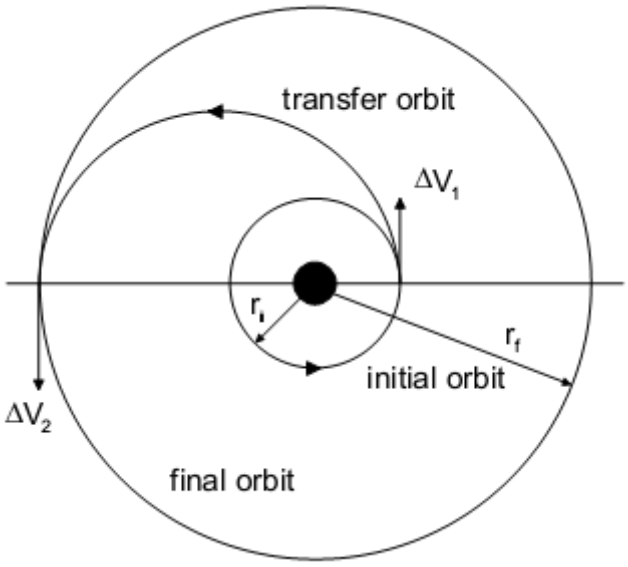
If the initial orbit has a radius larger than the final orbit, the same strategy can be followed but in this case, negative impulses will be required, first at apogee and then at perigee, to decelerate the satellite.

Whether moving into a higher or lower orbit, the time taken to transfer between the orbits is:

$$t_H = \frac{1}{2}\sqrt{\frac{4\pi^2 a_H^3}{\mu}} = \pi\sqrt{\frac{(r_1 + r_2)^3}{8\mu}}$$

Which is one half of the orbital period for the whole ellipse, where a_H is length of semi-major axis of the Hohmann transfer orbit.

The following diagram illustrates the geometry of the coplanar Hohmann transfer:



Example

A spacecraft is in a circular parking orbit with an altitude of 200 km. Calculate the velocity change required to perform a Hohmann transfer to a circular orbit at geosynchronous altitude.

Sol:

Given: rA = (6,378.14 + 200) × 1,000 = 6,578,140 m

rB = 42,164,170 m

a = (rA + rB) / 2

a = (6,578,140 + 42,164,170) / 2

a = 24,371,155 m

V1 = SQRT[GM / rA] = SQRT[3.986005×10^{14} / 6,578,140]

V1 = 7,784 m/s

V2 = SQRT[GM / rB] = SQRT[3.986005×10^{14} / 42,164,170]

V2 = 3,075 m/s

Vp = SQRT[GM × (2 / rA - 1 / atx)]

Vp = SQRT[3.986005×1014 × (2 / 6,578,140 - 1 / 24,371,155)]

Vp = 10,239 m/s

Vap = SQRT[GM × (2 / rB - 1 / atx)]

Vap = SQRT[3.986005×1014 × (2 / 42,164,170 - 1 / 24,371,155)]

Vap = 1,597 m/s

ΔV1 = Vp – V1 = 10,239 - 7,784

ΔV1 = 2,455 m/s

ΔV2 = V2 – Vap = 3,075 - 1,597

ΔV2 = 1,478 m/s

ΔVT = ΔV1 + ΔV2 = 2,455 + 1,478 , ΔVT = 3,933 m/s

Ex- For the geostationary transfer orbit we have r_2 = 42,164 km and e.g. r_1 = 6,678 km (altitude 300 km).

In the smaller circular orbit the speed is 7.73 km/s, in the larger one 3.07 km/s. In the elliptical orbit in between the speed varies from 10.15 km/s at the perigee to 1.61 km/s at the apogee.
The delta-v's are 10.15 − 7.73 = 2.42 and 3.07 − 1.61 = 1.46 km/s, together 3.88 km/s.

- Students need to validate the above exercise!

Exercise

A communication satellite was carried by the Space Shuttle into low earth orbit (LEO) at an altitude of 322 km and is to be transferred to a geostationary orbit (GEO) at 35, 860 km using a Hohmann transfer. Determine the characteristics of the transfer ellipse and the total Δv required.

From the inclination change equation, we see that if the angular change is equal to 60 degrees, the required change in velocity is equal to the current velocity.

Plane changes are very expensive in terms of the required change in velocity and resulting propellant consumption. To minimize this, we should change the plane at a point where the velocity of the satellite is a minimum: at apogee for an elliptical orbit. In some cases, it may even be cheaper to boost the satellite into a higher orbit, change the orbit plane at apogee, and return the satellite to its original orbit.

Typically, orbital transfers require changes in both the size and the plane of the orbit, such as transferring from an inclined parking orbit at low altitude to a zero-inclination orbit at geosynchronous altitude. We can do this transfer in two steps: a Hohmann transfer to change the size of the orbit and a simple plane change to make the orbit equatorial.

A more efficient method (less total change in velocity) would be to combine the plane change with the tangential burn at apogee of the transfer orbit. As we must change both the magnitude and direction of the velocity vector, we can find the required change in velocity using:

$$\Delta V = \sqrt{V_i^2 + V_f^2 - 2V_i V_f \cos\theta}$$

where Vi is the initial velocity, V_f is the final velocity, and $\vartheta$is the angle change required.

Example

A satellite is in a parking orbit with an altitude of 200 km and an inclination of 28 degrees. Calculate the total velocity change required to transfer the satellite to a zero-inclination geosynchronous orbit using a Hohmann transfer with a combined plane change at apogee.

Given: r1 = (6,378.14 + 200) × 1,000 = 6,578,140 m
r2 = 42,164,170 m
ϑ = 28 degrees

V2 = 3,075 m/s
V1 = 1,597 m/s
ΔV1 = 2,455 m/s (last Example)

ΔV2 = SQRT[VtxB2 + VfB2 - 2 × VtxB × VfB × cos θ]
ΔV2 = SQRT[1,5972 + 3,0752 - 2 × 1,597 × 3,075 × cos(28)]
ΔV2 = 1,826 m/s

ΔVT = ΔV1 + ΔV2 = 2,455 + 1,826
ΔVT = 4,281 m/s

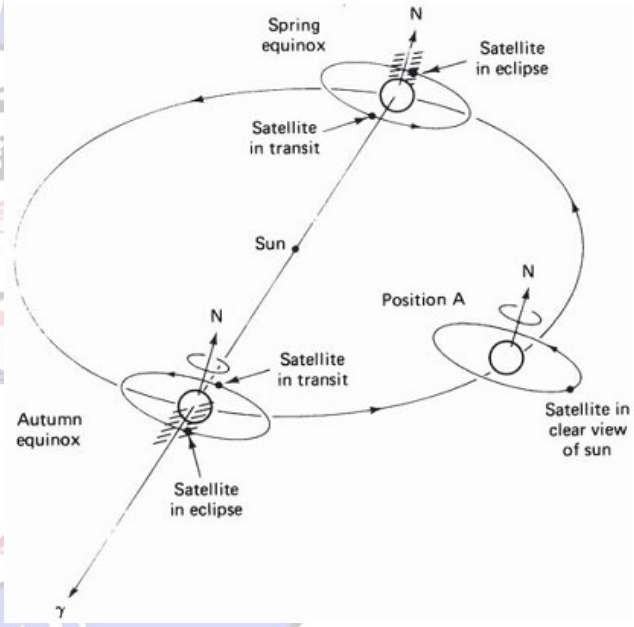
As can be seen from above, a small plane change can be combined with an altitude change for almost no cost in ΔV or propellant. Consequently, in practice, geosynchronous transfer is done with a small plane change at perigee and most of the plane change at apogee.

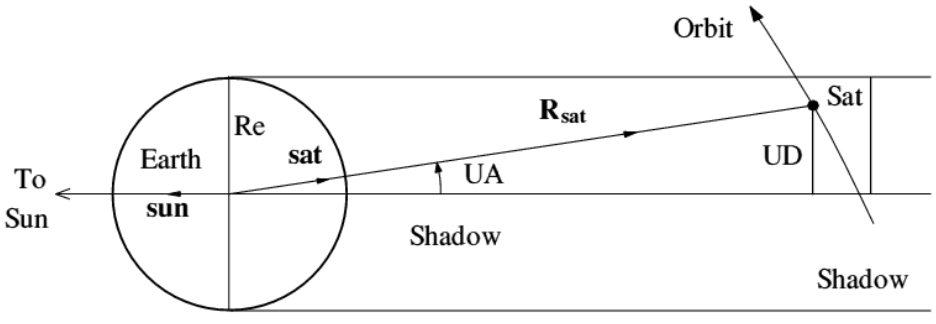
WEEK 5

Solar Eclipse

If the earth's equatorial plane coincided with the plane of the earth's orbit around the sun (the ecliptic plane), geostationary satellites would be eclipsed by the earth once each day. As it is, the equatorial plane is tilted at an angle of 23.4°to the ecliptic plane, and this keeps the satellite in full view of the sun for most days of the year, as illustrated by position A in Fig. Around the spring and autumnal equinoxes, when the sun is crossing the equator, the satellite does pass into the earth's shadow at certain periods, these being periods of eclipse as illustrated in Fig. The spring equinox is the first day of spring, and the autumnal equinox is the first day of autumn.

Eclipses begin 23 days before equinox and end 23 days after equinox. The eclipse lasts about 10 min at the beginning and end of the eclipse period and increases to a maximum duration of about 72 min at full eclipse. During an eclipse, the solar cells do not function, and operating power must be supplied from batteries. Figure (a) shows eclipse time as a function of height.





ECLIPSE GEOMETRY

Ex. A LEO satellite with 600 km height, calculate eclipse period & frequency. Take the Earth radius as 6400 Km.

According to the geometry,

$$\theta = \sin^{-1}\frac{6400}{7000} = 66$$

$2\theta/360=0.36$ which means 36% of its period the satellite will be in eclipse

We previously knew - from Kepler's rule - that, $a^3=\mu/n^2$

n=0.001078 rad/s

$p=2\pi/n$

p=5828.519 sec. = 97.142 min = 1.619 hrs.

This means that the satellite is orbiting the earth more than 14 times a day, so 14 eclipse per day take places.

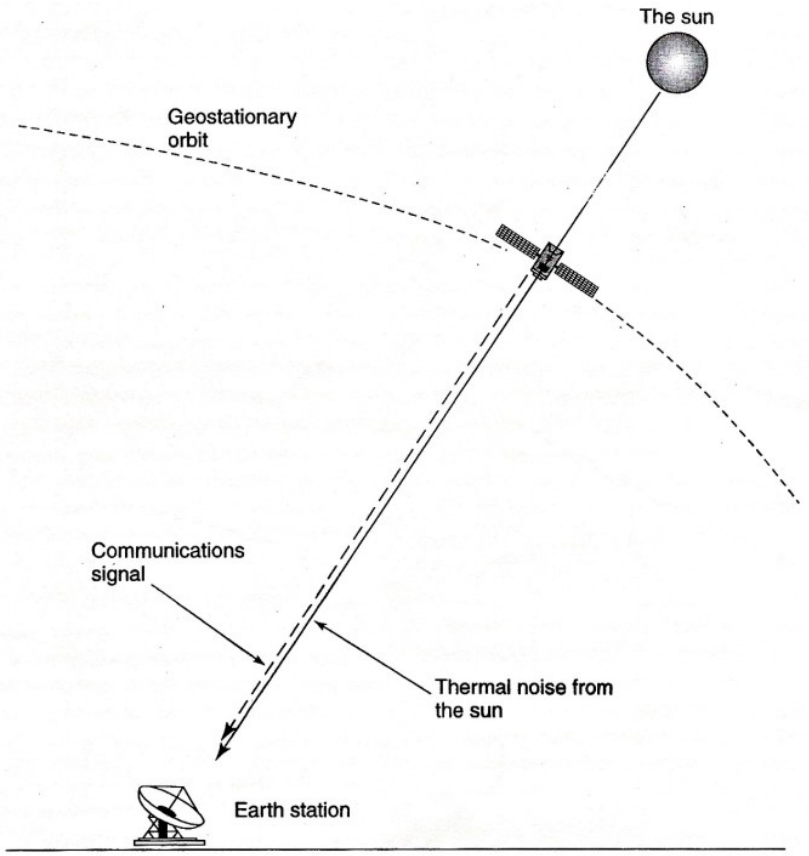
Now 36 % eclipse for a period of 97.142 min=~100min will be a 36min eclipse period.

Exercise1: Check the eclipse period & frequency for LEO satellite with 200 km height.

Exercise2: Check the eclipse period & frequency for GEO satellite in minutes.

Sun Transit Outage

During the equinox periods, not only does the satellite pass through the earth's shadow on the "dark" side of the earth, but the orbit of the satellite will also pass directly in front of the sun on the sunlit side of the earth (Figure 2.23). The sun is a "hot" microwave source with an equivalent temperature of about 6000 to 10,000 K, depending on the time within the 11-year sunspot cycle, at the frequencies used by communications satellites (4 to 50 GHz). The earth station antenna will therefore receive not only the signal from the satellite but also the noise temperature transmitted by the sun. The added noise temperature will cause the fade margin of the receiver to be exceeded and an outage will occur. These outages may be precisely predicted. For satellite system operators with more



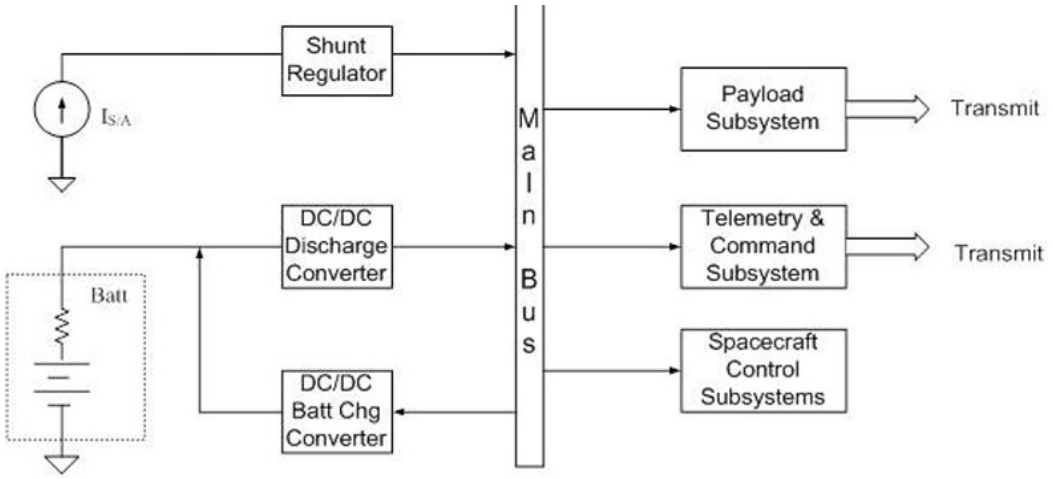
SATELLITE POWER SYSTEMS

The primary electrical power for operating the electronic equipment is obtained from solar cells. Individual cells can generate only small amounts of power, and therefore, arrays of cells in series-parallel connection are required.

Higher powers can be achieved with solar panels arranged in the form of rectangular solar sails. Solar sails must be folded during the launch phase and extended when in geostationary orbit. The solar cells are folded up on each side, and when fully extended, they stretch to its full area. The full complement of solar cells is exposed to the sunlight, and the sails are arranged to rotate to track the sun, so they are capable of greater power output than cylindrical arrays having a comparable number of cells. The HS 601 model can be designed to provide dc power from 2 to 6 kW.

The satellite power system is responsible for supplying electric energy to various satellite subsystems. Needless to mention that this subsystem has to meet stringent mass and volume limitation.

A typical power system is shown below:



Typical satellite power system

DC/DC Battery Charge Convertor -Battery Charge Regulator- (BCR)

BCR is a module responsible for charging the batteries during sun light. BCR monitors the charge state of the batteries and determines charging current accordingly; it should work at high efficiency.

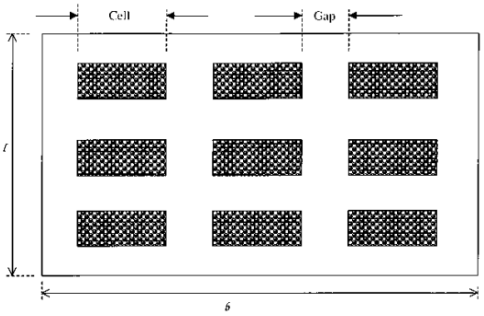
DC/DC Battery Discharge Convertor -Battery Discharge Regulator- (BDR)

During eclipse, when the solar panel stop generating electric power, and the spacecraft electric bus is powered by the rechargeable batteries, the BDR controls the output voltage from the batteries and provides the necessary bus voltage.

A key constraint on satellites is power. Solar cells can power the satellite transponder. Due to the inherent risk of nuclear fuel, solar energy power becomes attractive.

Above Earth's atmosphere, the average solar flux that falls on a spacecraft solar panel is about 1353 W/m^2 (the solar constant). If we are able to harness the sun's energy, we will be able to develop a power source for the satellite; the power source may be defined as:

$$P_s = k_s A_a \eta_{sc} \eta_e cos\theta$$



P_s : effective solar system power (w).
K_s : solar constant .
A_a: panel area.
η_e: Ratio of effective solar cell area to the panel area (80~98) %
η_{sc}: solar cells' conversion efficiency; depends on the material used, for example:

- η_{sc}: ~26% for gallium arsenide (GaAs) semiconductor.
- η_{sc}: (10~14) % for silicon semiconductor.

θ : inclination angle of the panel with respect to the sun light.

- ☒ Solar cell efficiency is reduced by 30% when operated in outer space for long period (~7 years) due to the exposure to the nuclear radiation.

During eclipses, solar cells are inactive. To keep the communication system operating, backup batteries make up the power system.

Ex. A system requires 150W of continuous power for a communication satellite. Design a power source that meets this requirement, you have only gallium arsenide with an unused area of 15%.

There are several types of solar panel configuration exist, the most commonly used can be order into the following categories:

i- A spherical (or nearly spherical) satellite can be designed where solar panel covers the satellite skin. In this case, no orientation (attitude) on the satellite is required. The area exposed to the sun is given by:

$$A_{Sun} = \frac{1}{4} A_{total}$$

where A_{Sun} is the effective area illuminated by Sun.
A_{total} is the total spherical area.

ii- Cylindrical satellite with spin stabilization. Solar panels cover the side skin of the satellite. In this case the effective area exposed to the Sun is given by:

$$A_{sun} = \frac{1}{\pi} A_{total} \cos\delta$$

Where δ is the angle made by the sun with respect to the plane perpendicular to the spin axis.

iii- Body stabilized satellite: In this class deployable solar panel can be used, solar panels normally are design to track the Sun in its motion. However, variation of declination of the sun is normally not compensated for the sake of simplicity. In this case:

$$A_{sun} = A_{total} \cos\theta$$

where θ is the angle between the sun elevation and the plane perpendicular to the rotation axis of the sun tracking motor.

Secondary Energy Source:

The secondary energy source store energy from the primary source & provide the stored energy when the primary stop functioning (e.g. during eclipse). Types of batteries used are Nickle-Cademium and Nickle-Hydrogen.

Their capacity is determined by:

i- Spacecraft load during the eclipse.
ii- Eclipse duration.
iii- Frequency of eclipse.

To compute the battery capacity, the following equation is used:

$$Battery\ Capacity(in\ A\ H) = \frac{W_{eclipse}}{V_B} \cdot \frac{T_{eclipse}}{DoD}$$

where:

$W_{eclipse}$: Maximum load during eclipse.

V_B: Battery Voltage

$T_{eclipse}$: Maximum duration of eclipse time.

DoD: Allowable depth of discharge.

For geo satellites, they encounter 84 eclipses per year & DoD is around 0.4 to 0.5.

For LEO the case is different, the satellite encounters around 15 eclipse per day with about 35 minute duration. This means the battery has more than 5000 charge/discharge per year, for this reason; batteries for LEO are allowed for (0.15~0.2) DoD.

Transponder

A transponder is the series of interconnected units which forms a single communications channel between the receive and transmit antennas in a communications satellite. Some of the units utilized by a transponder in a given channel may be common to a number of transponders. Thus, although reference may be made to a specific transponder, this must be thought of as an equipment channel rather than a single item of equipment.

The bandwidth allocated for C-band service is 500 MHz, and this is divided into sub-bands, one for each transponder. A typical transponder bandwidth is 36 MHz, and allowing for a 4-MHz guard band between transponders, 12 such transponders can be accommodated in the 500-MHz bandwidth. By making use of polarization isolation, this number can be doubled. Polarization isolation refers to the fact that carriers, which may be on the same frequency but with opposite senses of polarization, can be isolated from one another by receiving antennas matched to the incoming polarization. With linear polarization, vertically and horizontally polarized carriers can be separated in this way, and with circular polarization, left-hand circular and right-hand circular polarizations can be separated.

Because the carriers with opposite senses of polarization may overlap in frequency, this technique is referred to as *frequency Re-use*. Figure shows part of the frequency and polarization plan for a C-band communications satellite.

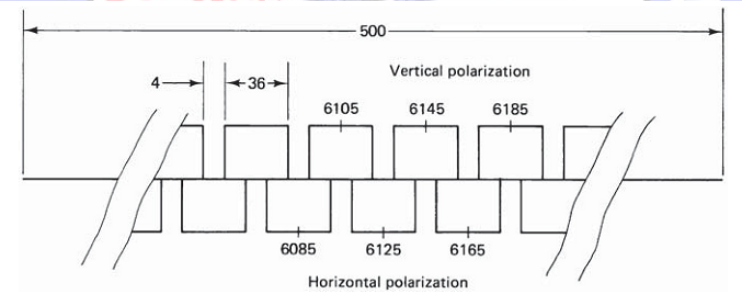
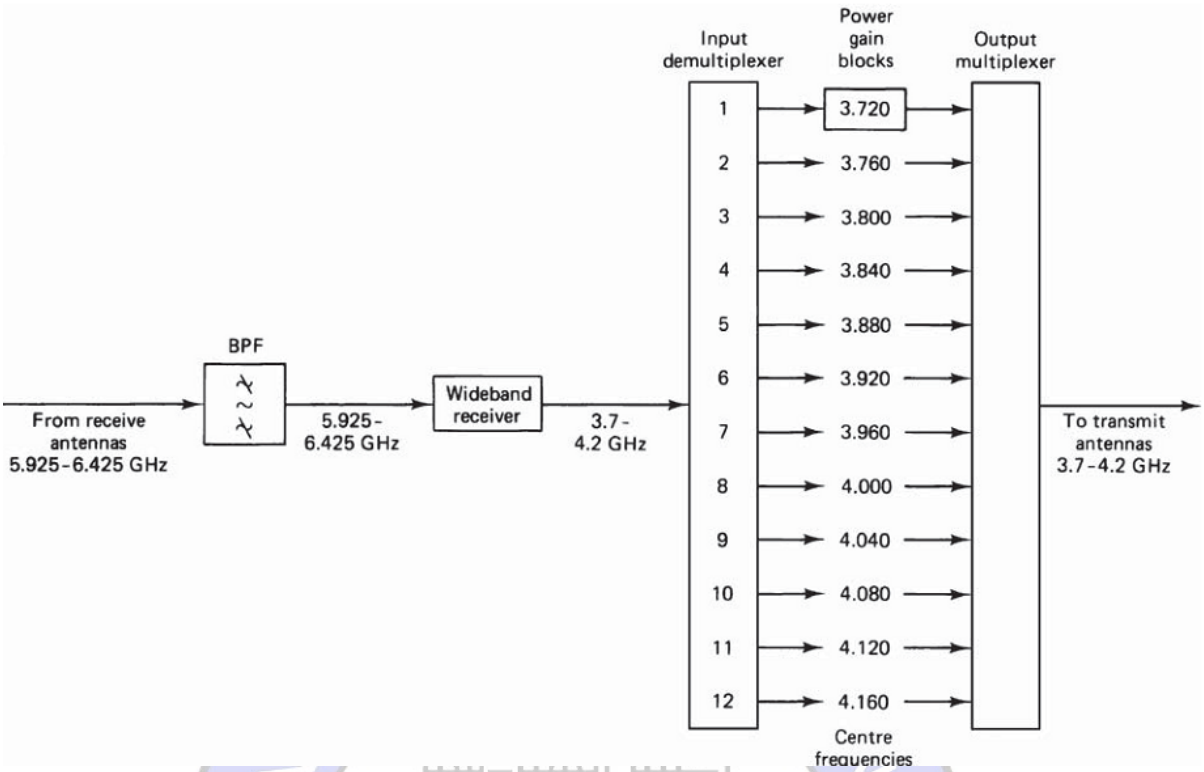
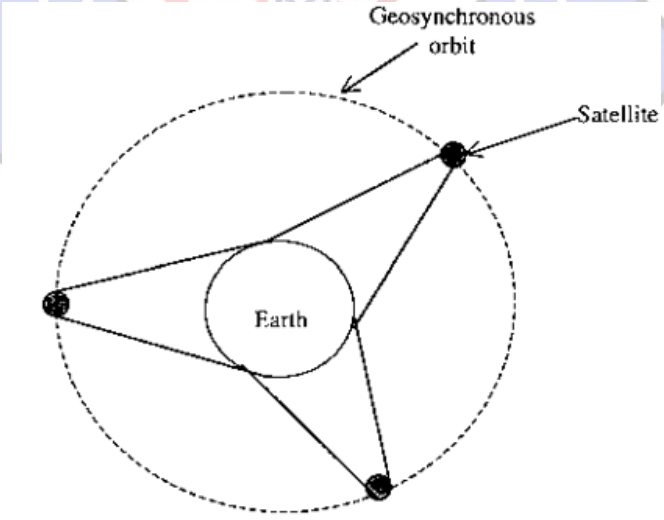


Fig.(4) Section of an uplink frequency and polarization plan. Numbers refer to frequency in megahertz.



COVERAGE AREA AND SATELLITE NETWORKS

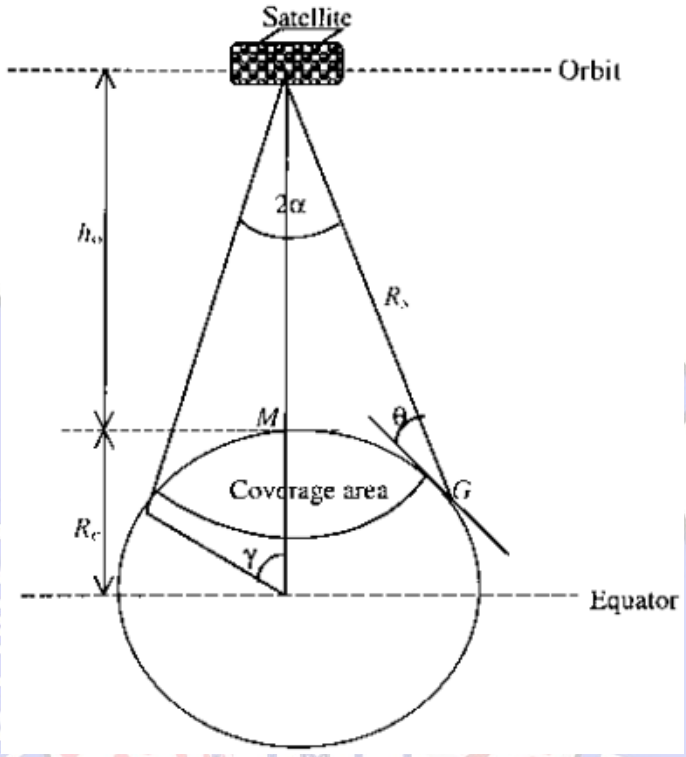
Clarke foresaw in his article that it would be possible to provide complete radio coverage of the world from just three satellites, provided that they could be precisely placed in geosynchronous orbit. Figure below demonstrates this.



Complete coverage of the earth's surface from three satellites.

The amount of coverage is an important feature in the design of earth observation satellites. Coverage depends on altitude and look angles of the equipment, among several factors.

To establish the geometric relationship of the coverage, we take a section of the satellites in the above figure as an illustration.



The maximum geometric coverage can be defined as the portion of the earth within a cone of the satellite at its apex, which is tangential to the earth's surface. Consider the angle of view from the satellite to the earth terminal as α; then the apex angle is 2α. The view angle has a mathematical physical function given by:

$$\alpha = \sin^{-1}\left(\frac{R_e}{h_0 + R_e}\right) = \sin^{-1}\left(\frac{R_e}{r}\right)$$

Exercise: Compute the minimum angle that's required by a satellite with 1000 km height to cover the maximum geometric coverage of the earth.

WEEK 6

PATH LOSS

The path loss elements include free-space loss, atmospheric losses due to gaseous and water vapor absorption, precipitation, fading loss due to multipath, and other miscellaneous effects based on frequency and the environment.

If the principal path is governed by free-space loss, it is calculated using the Friis free-space loss equation, which can be expressed as:

$$L = G_T G_R \left(\frac{\lambda}{4\pi d} \right)^2$$

The antenna gain accounts for the antenna directivity and efficiency, while the inverse distance squared term accounts for the spherical wave-front (geometric) spreading.

The Friis free-space loss equation can be expressed in dB as:

$$L_{dB} = -G_{TdB} - G_{RdB} - 20\log(\lambda) + 20\log(d) + 22$$

where a negative sign has been included so that the value of L_{dB} is in fact a loss.

In many applications, the antenna gains are excluded from the path loss expression, in which case the free-space loss is:

$$L_{FSL\,dB} = -20\log\left(\frac{\lambda}{4\pi d} \right)$$

The link margin is obtained by comparing the expected received signal strength to the receiver sensitivity or threshold.

The link margin is a measure of how much margin there is in the communications link between the operating point and the point where the link can no longer be used. The link margin can be found using:

$$Link\ margin\ =\ EIRP - L_{path} + G_R - TH_{Rx}$$

where

EIRP is the Effective Isotropic Radiated Power in dBm

L_{Path} is the total path loss, including miscellaneous losses, reflections, and fade margins in dB

G_R is the receiver gain in dB

TH_{Rx} is the receiver threshold or the minimum received signal level that will provide reliable operation in dBm

Example/Consider a 100-m link that operates in free space at 10 GHz. Assume that the transmit power is 0.1 W, and both transmit and receive antennas have 5-dB gain. If the receiver threshold is -85 dBm, what is the available link margin?

Sol:

The first step is to compute the EIRP of the transmitter.

P_T =100 mW = 20 dBm

EIRP = $P_T + G_T$ = 20+5 = 25 dBm

$$L_{FSL\,dB} = -20\log\left(\frac{\lambda}{4\pi d}\right)$$

$$= -20\,log\left(\frac{0.3}{4\pi * 100}\right) = 92.44\,dB$$

The resulting link margin is:

$$M_L = 25 - 92.4 + 5 - (-85) = 22.6\,dB$$

RECEIVER GAIN

The receiver gain is equal to the antenna gain less any radome loss, cable or waveguide loss (receiver loss), polarization loss, and pointing loss.

$$G_{RdB} = G_{RxdB} - L_{RadomedB} - L_{WGdB} - L_{Pol} - L_{Pt}$$

The received signal level RSL is the EIRP minus the path loss, plus the receiver gain:

$$\mathrm{RSL} = \mathrm{EIRP} - \mathrm{PL} + G_{RdB}$$

The signal-to-noise ratio can then be computed using the RSL and the noise floor:

$$\mathrm{SNR} = \mathrm{RSL}_{\mathrm{dBm}} - N_{\mathrm{dBm}}$$

The allowable path loss is given by the sum (in dB) of the free-space loss, the fade margin (FM_{dB}), and any miscellaneous losses:

$$\mathrm{PL}_{dB} = \mathrm{FSL}_{dB} + \mathrm{FM}_{dB} + L_{miscdB}$$

NOISE

Receiver noise is caused by the thermal agitation of electrons, in the front end of the receiver. This noise is modeled as having a Gaussian amplitude distribution. In addition, it is modeled as having a flat (constant) power spectral density of $N_o/2$, meaning that the power density is the same over all frequencies of interest. Thus, the thermal noise usually characterized as "white". Finally, the thermal noise is additive in nature (i.e., independent of the actual signal involved). Therefore, thermal noise is modeled as additive white Gaussian noise (AWGN).

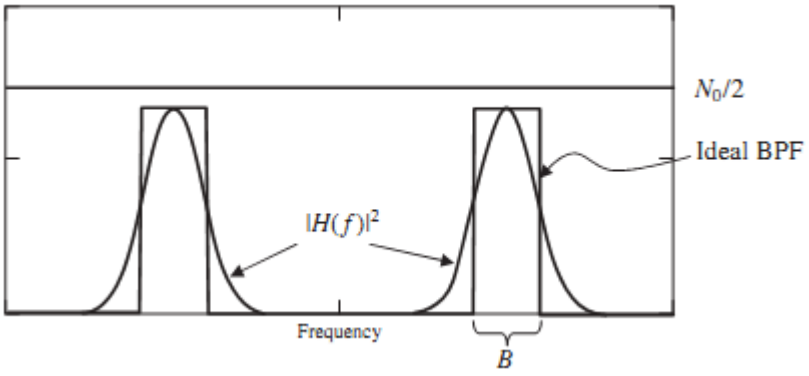
The theoretical thermal noise power available to an antenna is:

$$N = kT_oB$$

where:
$K = 1.38 * 10^{-23}\ J/K\ (Boltzman\ Constant)$
$T_o = 290\ K\ (Standard\ Noise\ Temprture)$
$B = Noise\ Bandwidth\ in\ (Hz)$

The noise-equivalent bandwidth is the bandwidth of an ideal band-pass filter that would pass the same amount of white noise power as the system being analyzed. In most modern digital communication systems, it is closely approximated by:

$$B_N = \frac{1}{T_s} \quad \text{Hz}$$



Real-world components are sources of additional noise so that the overall noise level is always greater than the theoretical value. For example, a receiver adds noise of its own by the very presence of resistive elements in its front end. Receiver noise can be characterized by either the effective noise temperature or the noise figure (they are related). The actual noise power is then:

$$N = kT_oB + kT_eB \quad \text{W}$$

$$N = kT_oBF \quad \text{W}$$

where F = (1 + T_e /T_0) is called the noise factor and T_e is the equivalent noise temperature of the receiver. This term, F, is generally referred to as the noise figure, when expressed in dB.

The computation of the total noise power can also be performed in dB:

$$N_{dBm} = -174\frac{dBm}{Hz} + 10\ log\ (B)dB.Hz + F_{dB}$$

The noise figure can also be defined as the ratio of the input signal-to-noise ratio to the output signal-to-noise ratio.

$$F = \frac{SNR_{in}}{SNR_{out}}$$

Where:

$$SNR_{in} = \frac{S}{KT_0B}$$

$$SNR_{out} = \frac{S}{K(T_0 + T_e)B}$$

By substitution:

$$F = 1 + \frac{T_e}{T_0}$$

Example:
Given a receiver for a 10-megasymbol-per-second signal, determine the available noise power present with the signal and find the noise figure of the receiver if the equivalent noise temperature of the receiver is 870 K.

Sol:

$$10\frac{\text{Ms}}{\text{s}} \equiv 10\ MHz = f$$

The noise figure is:

$$F = 1 + \frac{T_e}{T_0}$$

$$= 1 + \frac{870}{290} = 4 = 6\ dB$$

The noise power present with the received signal will be:

$$N = KT_0BF$$

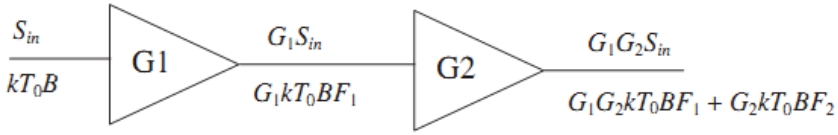
$$N = 10\log(1.38 * 10^{-23} * 290 * 10 * 10^6 * 4)$$

$$N = -128\ dBW$$

When there are several cascaded elements to be considered, the noise figure of the composite system can be readily determined. First, consider the case of two cascaded amplifiers as shown in the below figure. Assume the system is driven by a source whose impedance matches the first amplifier stage and whose noise level is:

$$N_{in} = kT_0B$$

The accompanying signal is S_{in}.



The output of the first stage will be:

$$S_1 = G_1S_{in} + G_1N_{in} + G_1kT_{e1}B$$

Or

$$S_1 = G_1S_{in} + G_1kT_0BF_1$$

where T_{e1} is the effective temperature of the first stage, referenced to the input.
The signal at the input to the second amplifier is the output of the first amplifier, S1. To determine the output of the second amplifier, we add the apparent noise of the second amplifier, which is characterized by $T_0 + T_{e2}$, and multiply by the gain. Thus the output of the second amplifier is:

$$S_2 = G_1G_2S_{in} + G_1G_2kT_0BF_1 + G_2kT_0BF_2$$

To find the noise figure of the combined system, the definition of the noise figure is applied:

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

$$F_{total} = \frac{S_{in}/kT_0B}{G_1G_2S_{in}/[G_1G_2kT_0B(F_1 + F_2/G_1)]}$$

$$F_{total} = F_1 + \frac{F_2}{G_1}$$

it can be shown that for any number of cascaded amplifiers, the composite noise figure is given by:

$$F_{total} = F_1 + \frac{F_2}{G_1} + \frac{F_3}{G_1G_2} + \frac{F_4}{G_1G_2G_3} + \ldots$$

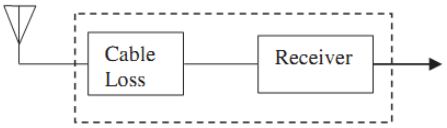
Based on the preceding analysis, it can be seen that the first stage of a well-designed system sets the noise floor and thus must have the best noise performance of all of the stages and relatively high gain.

Any losses present between the antenna output and the receiver input increase the overall noise figure of the system. For terrestrial systems, where the input temperature is assumed to be the standard noise temperature of 290 K, the new noise figure is computed by simply adding the losses to the receiver noise figure.

Example:

Consider the receiver system shown below:

Given that the cable loss is 7 dB and the noise equivalent temperature of the receiver is 630 K, what is the noise figure of the receiver system?



SOL:

The noise figure of the receiver alone is:

$$F = 1 + \frac{T_e}{T_0} = 5\ dB$$

The total noise figure of the receiver will be 7+5=12 dB.

Q1/ Consider a receiver with a 4-dB noise figure and a 2-MHz bandwidth. If a carrier-to-noise ratio of 15 dB is required for acceptable BER performance, what signal strength is required at the receiver?

Q2/ Given a receiver with a 100-kHz bandwidth and an effective noise temperature of 600 K, what is the noise power level at the input to the receiver?

Q3/ Consider a point-to-point communication link operating at 38 GHz, using 35-dB dish antennas at each end. The radome loss is 2 dB at each end, and a 10-dB fade margin is required to allow for rain fades. If the transmit power is -10 dBm, the receiver noise figure is 7 dB, the bandwidth is 50 MHz, and the link distance is 1.2 km, what is the signal-to-noise ratio at the receiver in dB? Assume that free-space loss applies.

Q4 / A communication link operating at 28 GHz, using 30-dB dish antennas at each end. A 10-dB fade margin is required to allow for rain fades. If the transmit power is 0 dBm, the receiver noise figure is 7 dB, the bandwidth is 50 MHz, and the required carrier-to-noise ratio for reliable operation is 12 dB, find the following (Assume that free-space loss applies):

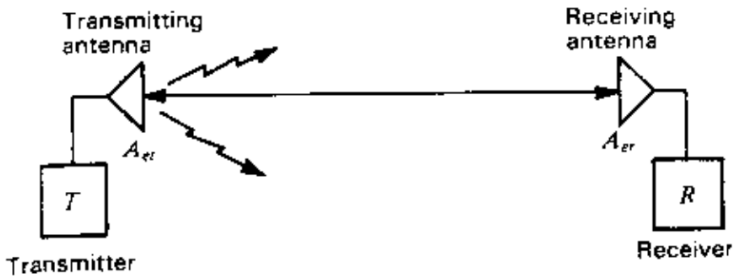
(a) The required received signal level. (b) The allowable path loss.

(c) The maximum allowable range.

WEEK 7

FRIIS TRANSMISSION FORMULA

The important *Friis transmission formula* published in 1946 by Harald T. Friis.



Referring to figure above, this formula gives the power received over a radio communication circuit. Let the transmitter T feed a power P_t to a transmitting antenna of effective aperture A_{et} . At a distance (r) a receiving antenna of effective aperture A_{er} intercepts some of the power radiated by the transmitting antenna and delivers it to the receiver R. Assuming for the moment that the transmitting antenna is isotropic, the power per unit area at the receiving antenna is:

$$S_r = \frac{P_t}{4\pi r^2}$$

& for an antenna with certain gain G_t :

$$S_r = \frac{P_t\, G_t}{4\pi r^2}$$

The power collected by the receiving antenna of effective aperture A_{er} is:

$$P_r = S_r A_{er} = \frac{P_t\, G_t\, A_{er}}{4\pi r^2} \quad (W)$$

The transmitting antenna gain can be expressed as:

$$G_t = \frac{4\pi\, A_{et}}{\lambda^2}$$

Substituting this gain in the equation yields the Friis equation:

$$P_r = \frac{P_t\, A_{et}\, A_{er}}{r^2\, \lambda^2} \quad (W)$$

where P_r = received power , W
P_t = power into transmitting antenna, W
A_{et} = effective aperture of transmitting antenna, m^2
A_{er} = effective aperture of receiving antenna, m^2
r = distance between antennas, m
λ = wavelength, m

Q/ What is the maximum power received at a distance of 0.5 km over a free-space 1-GHz circuit consisting of a transmitting antenna with a 25 dB gain and a receiving antenna with a 20-dB gain? The gain is with respect to a lossless isotropic source. The transmitting antenna input is 150 W.

Sol:

$$\lambda = c/f = 3\times10^8/10^9 = 0.3 \text{ m},$$

$$A_{et} = \frac{D_t\lambda^2}{4\pi}, \qquad A_{er} = \frac{D_r\lambda^2}{4\pi}$$

$$P_r = P_t\frac{A_{et}A_{er}}{r^2\lambda^2} = P_t\frac{D_t\lambda^2 D_r\lambda^2}{(4\pi)^2 r^2\lambda^2}$$

$$= 150\frac{316\times0.3^2\times100}{(4\pi)^2 500^2} = 0.0108 \text{ W} = 10.8 \text{ mW}$$

Q/ Design a two-way radio link to operate over earth-Mars distances with 2.5 GHz & 5 MHz bandwidth. A power of 10^{-19} W /Hz is to be delivered to the earth receiver and 10^{-17} W/Hz to the Mars receiver. The Mars antenna must be no larger than 3 m in diameter & Earth antenna of 100 m in diameter with 0.5 efficiency for both. Specify transmitter power at each end. Take earth-Mars distance as 6 light-minutes.

Sol:

$$\lambda = c/f = 3\times10^8/2.5\times10^9 = 0.12 \text{ m}$$

$$P_r(\text{earth}) = 10^{-19}\times5\times10^6 = 5\times10^{-13} \text{ W}$$

$$P_r(\text{Mars}) = 10^{-17}\times5\times10^6 = 5\times10^{-11} \text{ W}$$

$$A_r(\text{Mars}) = (1/2)\pi\,1.5^2 = 3.5 \text{ m}^2$$

$$P_t(\text{Mars}) = 1 \text{ kW}$$

$$A_r = (1/2)\pi\ 50^2 = 3927 \text{ m}^2$$

$$P_t(\text{earth}) = P_r(\text{Mars})\frac{r^2\lambda^2}{A_{et}(\text{earth})A_{et}(\text{Mars})}$$

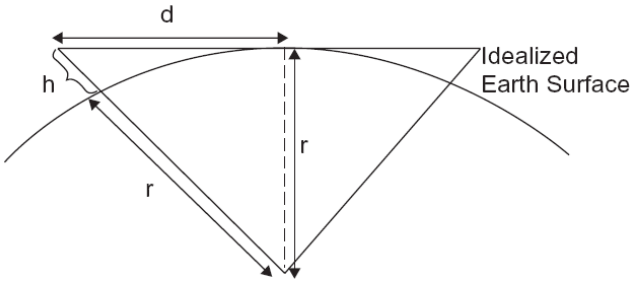
$$= 620 \text{ kW}$$

$$P_r(\text{earth}) = P_t(\text{Mars})\frac{A_{et}(\text{Mars})A_{er}(\text{earth})}{r^2\lambda^2} = 10^3\frac{3.5\times3930}{(360\times3\times10^8)^2 0.12^2} = 8\times10^{-14} \text{ W}$$

The required $5*10^{-13}$ W could be obtained by increasing the Mars transmitter power by a factor of 6.25

Line of Site & Earth Curvature

When considering line-of-sight (LOS) propagation, it may be necessary to consider the curvature of the earth (Figure below). The curvature of the earth is a fundamental geometric limit on LOS propagation. In particular, if the distance between the transmitter and receiver is large compared to the height of the antennas, then an LOS may not exist. The simplest model is to treat the earth as a sphere with a radius equivalent to the equatorial radius of the earth.



$$d^2 + r^2 = (r + h)^2$$
$$d^2 = (2r + h)h$$
$$d^2 = (2rh + h^2)$$

Since $rh \gg h^2, then$

$$d \cong \sqrt{2rh}$$

where:
r: earth radius
h is the antenna height from the surface of the earth.

The radius of the earth is approximately 6390 Km at the equator. The atmosphere typically bends horizontal RF waves downward due to the variation in atmospheric density with height, for now it is sufficient to note that an accepted means of correcting for this curvature is to use the "4/3 earth approximation," which consists of scaling the earth's radius by 4/3.

The effect of wave bending is taken into account by using the so-called 4/3's-earth model. The model consists of simply replacing the earth's radius by 4/3 of its actual value in horizon calculations. The 4/3's factor is based on an ITU reference standard atmosphere at sea level and as such represents an average. While not universal, the 4/3's-earth model is very widely used.
Thus:

$$d \cong \sqrt{2\frac{4}{3}rh} \quad or \quad d \cong \sqrt{\frac{8}{3}rh}$$

Exercise:
Given a point-to-point link with one end mounted on a 30m tower and the other on a 15m tower, what is the maximum possible (LOS) link distance?

WEEK 8

RADAR EQUATION

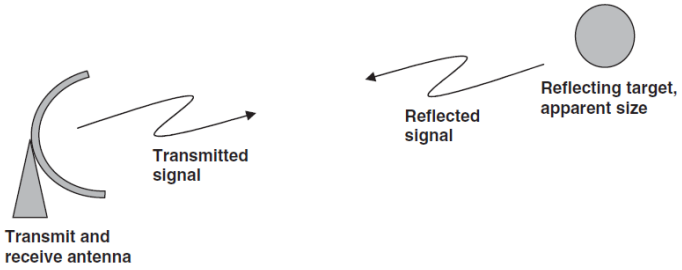
RADAR is an acronym for **RA**dio **D**etection **A**nd **R**anging.

Early radar was designed to detect large objects at a distance, such as aircraft or vessels.

As development continued, split antenna apertures were used to sense the direction of the returning wavefront (mono-pulse radar), providing greater angular accuracy. Another key development was the use of filter banks to detect the Doppler shift on the return signal which provided the operator with target velocity information. Modern imaging radars can also provide information on target size and sometimes target identification by using pulse compression and other range resolution techniques as well as synthetic aperture processing.
A radar "target" may be an aircraft, vehicle, vessel, weather front, or terrain feature, depending upon the application.

Consider the radar configuration shown in figure, where, the transmitter and receiver are co-located and share one antenna (monostatic radar) and the received signal is a reflection from a distant object. A pulse of RF energy is transmitted and then the receiver listens for the return, similar to sonar. The strength of the return signal depends upon the transmitted power, the distance to the reflecting target, and its electrical size or reflectivity.

The radar receiver determines the distance to the target by measuring the time delay between transmission and reception of the reflected pulse. By measuring the phase front (angle of arrival) of the returning signal, the radar can also estimate the location of the target in azimuth and elevation within its beamwidth



The radar range equation is used to determine the received radar signal strength, or signal-to-noise ratio. It can be used to estimate the maximum distance at which a target of a given size can be detected by the radar. The power density at a distance, *d*, is given by:

$$S = \frac{EIRP}{4\pi d^2}\ W/m^2$$

EIRP is (**E**ffective **I**sotropic **R**adiated **P**ower),

$$EIRP = P_T G_T$$

The power available at the output of an antenna in this power density field is the product of the power density at that point and the antenna's effective area :

$$P = P_T \frac{G_T}{4\pi d^2}.A_e \quad Watts$$

As we know that:

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

Then

$$P_R = \frac{P_T G_T G_R\, \lambda^2}{(4\pi)^2 d^2} \quad Watts$$

Instead of a receive antenna effective area, in radar, the return signal strength is determined by the radar cross section (RCS) of the "target." The RCS is a measure of the electrical or reflective area of a reflector. It may or may not correlate with the physical size of the object. It is a function of the object's shape, composition, orientation, and possibly size. The RCS is expressed in m^2, or dBsm (10log(RCS in m^2)) and the symbol for target RCS is σ. Since the RCS is an intrinsic property of the target or reflector, it does not depend on the range or distance from which it is viewed. It may, however, vary with the frequency and the polarization of the radar wave and with the aspect angle from which it is viewed.

The strength of the reflected signal is determined from the power density at the target by replacing the receive antenna capture area with the RCS.

$$P_{ref} = \frac{P_T G_T}{4\pi d^2}.\sigma_t \quad Watts$$

The power density back at the radar receiver from the reflected signal is:

$$S_R = \frac{P_T G_T}{4\pi d^2}.\frac{\sigma_t}{4\pi d^2} \qquad Watts/m^2$$

When multiplied by the effective area of the radar antenna, this becomes the received power.

$$P_R = \frac{P_T G_T \, \sigma_t \, A_e}{(4\pi)^2 d^4} \quad Watts$$

the expression for the received power can be written in terms of the antenna gains.

$$P_R = \frac{P_T G_T G_R \, \sigma_t \, \lambda^2}{(4\pi)^3 d^4} \quad Watts$$

For a monostatic radar, $G_T = G_R$.

$$P_R = \frac{P_T G^2 \, \sigma_t \, \lambda^2}{(4\pi)^3 d^4}$$

Using the minimum detectable receive power, one can solve the radar equation for d and find the maximum distance at which detection is possible. In radar, it is customary to use R for range instead of d for distance, so:

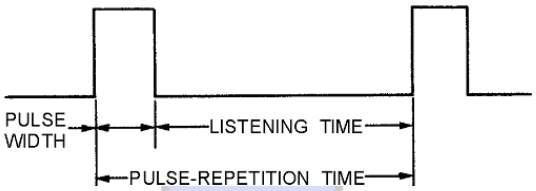
$$R_{max} = \sqrt[4]{\frac{P_T G^2 \sigma_t \, \lambda^2}{P_{Rmin}(4\pi)^3}} \quad m$$

This is called the radar range equation. It provides an estimate of the maximum radar detection range for a given target RCS. The maximum range is proportional to the transmit power, antenna gain, and target RCS and inversely proportional to the minimum required receive power and the radar frequency.

WEEK 9

RANGE AMBIGUITY

Radar systems radiate each pulse at the carrier frequency during transmit time, wait for returning echoes during listening or rest time, and then radiate a second pulse, as shown in figure below. The number of pulses radiated in one second is called the pulse-repetition frequency (prf) , pulse repetition time $prt = 1/prf$.



The radar timing system must be reset to zero each time a pulse is radiated. This is to ensure that the range detected is measured from time zero each time. The prt of the radar becomes important in maximum range determination because target return times that exceed the prt of the radar system appear at incorrect locations (ranges) on the radar screen. Returns that appear at these incorrect ranges are referred to as AMBIGUOUS RETURNS or SECOND-SWEEP ECHOES.

The maximum unambiguous range R_{unamb} is:

$$R_{unamb} = \frac{c}{2} * prt$$

Example:

A radar system with a 1 millisecond prt, scanning two targets A & B.

For target A, the pulse travels round trip in 0.5 millisecond, which equals to a target range of 75,000 meter.

Since 0.5 millisecond is less than 1 millisecond, displaying a correct range is no problem. However, target B is 180,000 meter distant from the radar system. In this case, total pulse travel time is 1.2 milliseconds and exceeds the prt limitation of 1 millisecond for this radar. While the first transmitted pulse is traveling to and returning from target B, a second pulse is transmitted and the radar system is reset to 0 again. The first pulse from target B continues its journey back to the radar system, but arrives during the timing period for the second pulse. This results in an inaccurate reading. In this case, the first return pulse from target B arrives 0.2 millisecond into the second timing period. This results in a range of 30,000 meter instead of the actual 180,000 meter.

RADAR TYPES

Radar systems are often divided into operational categories based on energy transmission methods to continuous wave (cw) and pulse modulation (pm).

The CONTINUOUS WAVE (cw) : transmits a constant frequency and detects moving targets by detecting the change in frequency caused by electromagnetic energy reflecting from a moving target. This change in frequency is called the DOPPLER SHIFT or DOPPLER EFFECT.

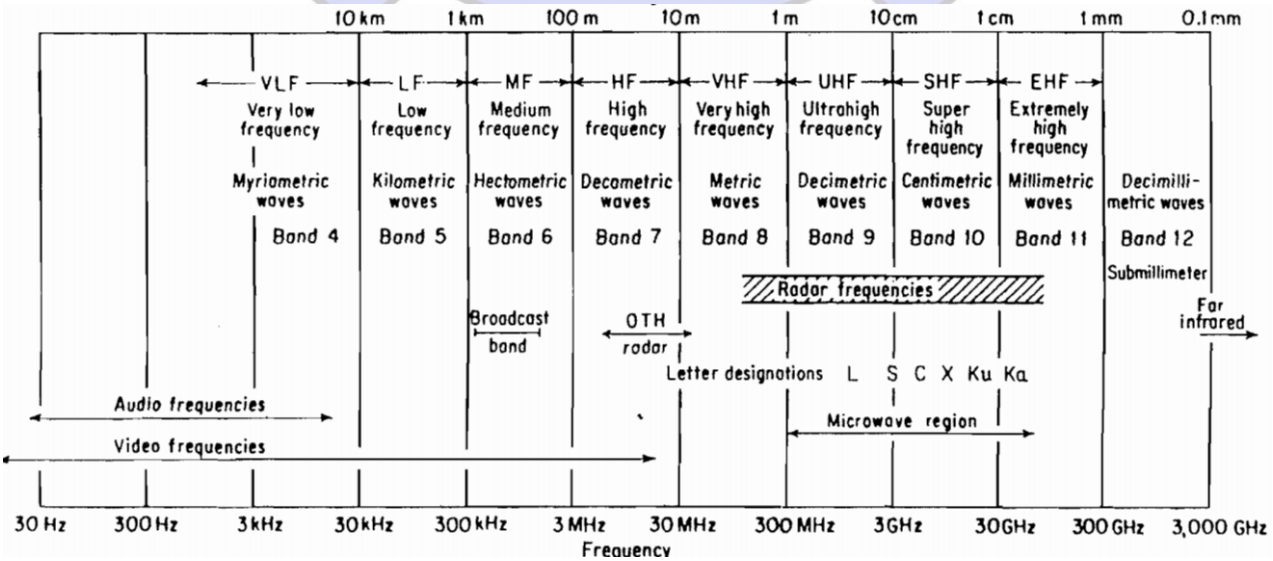
The PULSE-MODULATION (pm) : uses short pulses of energy and relatively long listening times to accurately determine target range. Since this method does not depend on signal frequency or target motion, it has an advantage over cw radar. It is the most common type of radar.

Radar systems are also classified by function. SEARCH RADAR continuously scans a volume of space and provides initial detection of all targets. TRACK RADAR provides continuous range, bearing, and elevation data on one or more specific targets. Most radar systems are variations of these two types.

RADAR FREQUENCIES

Conventional radars generally have been operated at frequencies extending from about 220 MHz to 35 GHz, a spread of more than seven octaves. These are not necessarily the limits, since radars can be, and have been operated at frequencies outside either end of this range.

Skywave HF over-the-horizon (OTH) radar might be at frequencies as low as 4 or 5 MHz, and groundwave HF radars as low as 2 MHz. At the other end of the spectrum, millimeter radars have operated at 94 GHz. Laser radars operate at even higher frequencies.



Q1/ A monostatic radar is employed to detect a target of a typical RCS of 4 m^2. Its operational frequency is 10 GHz, the transmitted power is 1000 W. If its antenna gain is 30 dBi, find the operational range of this radar.

Q2/ Consider a radar system with the following parameters:

f = 2 GHz Transmitter frequency
PT = 1 W Peak transmitter power (PT = 0 dBW)
GT = GR = 18 dB Antenna gains
R = 2 km Range to target
B = 50 kHz Receiver bandwidth
F = 5 dB Receiver noise figure
$\sigma_t = 1m^2$ Target radar cross section

What is the SNR at the receiver?

Q3/ Consider a radar system with

$P_T = 1{,}000{,}000$ W
$G_{ANT} = 28$ dB
$\tau = 100\,\mu$s
$T_{eff} = 200$ K
$f = 10$ GHz

What is the Noise figure of the system & SNR of the return from a 1-m^2 target at 20 km?

Q4 / Consider a radar system with the following parameters:

Transmit power Pt = 10.0 kW
Antenna gain G = 25 dB
Frequency f = 10 GHz

What is the received signal level from a 13-dBsm target at 10 km?

SPREAD-SPECTRUM MODULATION

This chapter introduces a modulation technique called spread-spectrum modulation, which is radically different from the modulation techniques that are covered in preceding chapters. In spread-spectrum modulation, channel bandwidth and transmit power are sacrificed for the sake of secure communications.

Specifically, we cover the following topics:

- *Spreading sequences in the form of pseudo-noise sequences, their properties, and methods of generation.*
- *The basic notion of spread-spectrum modulation.*
- *The two commonly used types of spread-spectrum modulation: direct sequence and frequency hopping.*

The material presented in this chapter is basic to wireless communications using code-division multiple access, which is covered in Chapter 8.

7.1 *Introduction*

A major issue of concern in the study of digital communications as considered in Chapters 4, 5, and 6 is that of providing for the efficient use of bandwidth and power. Notwithstanding the importance of these two primary communication resources, there are situations where it is necessary to sacrifice this efficiency in order to meet certain other design objectives. For example, the system may be required to provide a form of *secure* communication in a *hostile* environment such that the transmitted signal is not easily detected or recognized by unwanted listeners. This requirement is catered to by a class of signaling techniques known collectively as *spread-spectrum modulation.*

The primary advantage of a spread-spectrum communication system is its ability to reject *interference* whether it be the *unintentional* interference by another user simultaneously attempting to transmit through the channel, or the *intentional* interference by a hostile transmitter attempting to jam the transmission.

The definition of spread-spectrum modulation[1] may be stated in two parts:

1. Spread spectrum is a means of transmission in which the data sequence occupies a bandwidth in excess of the minimum bandwidth necessary to send it.
2. The spectrum spreading is accomplished before transmission through the use of a code that is independent of the data sequence. The same code is used in the receiver

(operating in synchronism with the transmitter) to despread the received signal so that the original data sequence may be recovered.

Although standard modulation techniques such as frequency modulation and pulse-code modulation do satisfy part 1 of this definition, they are not spread-spectrum techniques because they do not satisfy part 2 of the definition.

Spread-spectrum modulation was originally developed for military applications, where resistance to jamming (interference) is of major concern. However, there are civilian applications that also benefit from the unique characteristics of spread-spectrum modulation. For example, it can be used to provide *multipath rejection* in a ground-based mobile radio environment. Yet another application is in *multiple-access* communications in which a number of independent users are required to share a common channel without an external synchronizing mechanism; here, for example, we may mention a ground-based radio environment involving mobile vehicles that must communicate with a central station. More is said about this latter application in Chapter 8.

In this chapter, we discuss principles of spread-spectrum modulation, with emphasis on direct-sequence and frequency-hopping techniques. In a *direct-sequence spread-spectrum* technique, two stages of modulation are used. First, the incoming data sequence is used to modulate a wideband code. This code transforms the narrowband data sequence into a noiselike wideband signal. The resulting wideband signal undergoes a second modulation using a phase-shift keying technique. In a *frequency-hop spread-spectrum* technique, on the other hand, the spectrum of a data-modulated carrier is widened by changing the carrier frequency in a pseudo-random manner. For their operation, both of these techniques rely on the availability of a noiselike spreading code called a *pseudo-random* or *pseudo-noise sequence*. Since such a sequence is basic to the operation of spread-spectrum modulation, it is logical that we begin our study by describing the generation and properties of pseudo-noise sequences.

7.2 Pseudo-Noise Sequences

A *pseudo-noise (PN) sequence* is a periodic binary sequence with a noiselike waveform that is usually generated by means of a *feedback shift register*, a general block diagram of which is shown in Figure 7.1. A feedback shift register consists of an ordinary *shift register* made up of m flip-flops (two-state memory stages) and a *logic circuit* that are interconnected to form a multiloop *feedback* circuit. The flip-flops in the shift register are regulated by a single timing *clock*. At each pulse (tick) of the clock, the *state* of each flip-flop is shifted to the next one down the line. With each clock pulse the logic circuit computes a

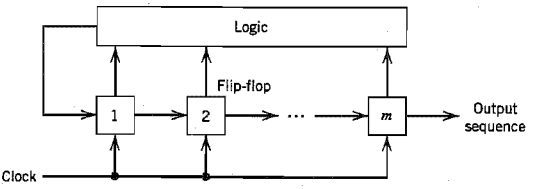


FIGURE 7.1 Feedback shift register.

Boolean function of the states of the flip-flops. The result is then fed back as the input to the first flip-flop, thereby preventing the shift register from emptying. The PN sequence so generated is determined by the length m of the shift register, its initial state, and the feedback logic.

Let $s_j(k)$ denote the state of the jth flip-flop after the kth clock pulse; this state may be represented by symbol 0 or 1. The state of the shift register after the kth clock pulse is then defined by the set $\{s_1(k), s_2(k), \ldots, s_m(k)\}$, where $k \geq 0$. For the initial state, k is zero. From the definition of a shift register, we have

$$s_j(k+1) = s_{j-1}(k), \qquad \begin{cases} k \geq 0 \\ 1 \leq j \leq m \end{cases} \tag{7.1}$$

where $s_0(k)$ is the input applied to the first flip-flop after the kth clock pulse. According to the configuration described in Figure 7.1, $s_0(k)$ is a Boolean function of the individual states $s_1(k), s_2(k), \ldots, s_m(k)$. For a specified length m, this Boolean function uniquely determines the subsequent sequence of states and therefore the PN sequence produced at the output of the final flip-flop in the shift register. With a total number of m flip-flops, the number of possible states of the shift register is at most 2^m. It follows therefore that the PN sequence generated by a feedback shift register must eventually become *periodic* with a period of at most 2^m.

A feedback shift register is said to be *linear* when the feedback logic consists entirely of *modulo-2 adders*. In such a case, the *zero state* (e.g., the state for which all the flip-flops are in state 0) is *not* permitted. We say so because for a zero state, the input $s_0(k)$ produced by the feedback logic would be 0, the shift register would then continue to remain in the zero state, and the output would therefore consist entirely of 0s. Consequently, the period of a PN sequence produced by a linear feedback shift register with m flip-flops cannot exceed $2^m - 1$. When the period is exactly $2^m - 1$, the PN sequence is called a *maximal-length-sequence* or simply *m-sequence*.

EXAMPLE 7.1

Consider the linear feedback shift register shown in Figure 7.2, involving three flip-flops. The input s_0 applied to the first flip-flop is equal to the modulo-2 sum of s_1 and s_3. It is assumed that the initial state of the shift register is 100 (reading the contents of the three flip-flops from left to right). Then, the succession of states will be as follows:

$$100, 110, 111, 011, 101, 010, 001, 100, \ldots.$$

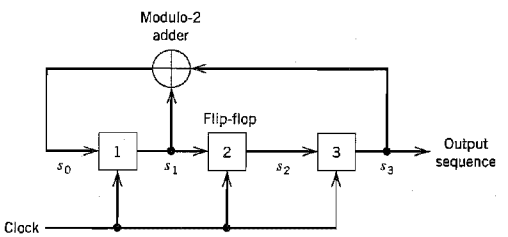


FIGURE 7.2 Maximal-length sequence generator for $m = 3$.

The output sequence (the last position of each state of the shift register) is therefore

$$00111010\ldots$$

which repeats itself with period $2^3 - 1 = 7$.

Note that the choice of 100 as the initial state is arbitrary. Any of the other six permissible states could serve equally well as an initial state. The resulting output sequence would then simply experience a cyclic shift. ◀

■ PROPERTIES OF MAXIMAL-LENGTH SEQUENCES[2]

Maximal-length sequences have many of the properties possessed by a truly *random binary sequence*. A random binary sequence is a sequence in which the presence of binary symbol 1 or 0 is equally probable. Some properties of maximal-length sequences are as follows:

1. In each period of a maximal-length sequence, the number of 1s is always one more than the number of 0s. This property is called the *balance property*.

2. Among the runs of 1s and of 0s in each period of a maximal-length sequence, one-half the runs of each kind are of length one, one-fourth are of length two, one-eighth are of length three, and so on as long as these fractions represent meaningful numbers of runs. This property is called the *run property*. By a "run" we mean a subsequence of identical symbols (1s or 0s) within one period of the sequence. The length of this subsequence is the length of the run. For a maximal-length sequence generated by a linear feedback shift register of length m, the total number of runs is $(N + 1)/2$, where $N = 2^m - 1$.

3. The autocorrelation function of a maximal-length sequence is periodic and binary-valued. This property is called the *correlation property*.

The period of a maximum-length sequence is defined by

$$N = 2^m - 1 \tag{7.2}$$

where m is the length of the shift register. Let binary symbols 0 and 1 of the sequence be denoted by the levels -1 and $+1$, respectively. Let $c(t)$ denote the resulting waveform of the maximal-length sequence, as illustrated in Figure 7.3*a* for $N = 7$. The period of the waveform $c(t)$ is (based on terminology used in subsequent sections)

$$T_b = NT_c \tag{7.3}$$

where T_c is the duration assigned to symbol 1 or 0 in the maximal-length sequence. By definition, the autocorrelation function of a periodic signal $c(t)$ of period T_b is

$$R_c(\tau) = \frac{1}{T_b}\int_{-T_b/2}^{T_b/2} c(t)c(t-\tau)\,dt \tag{7.4}$$

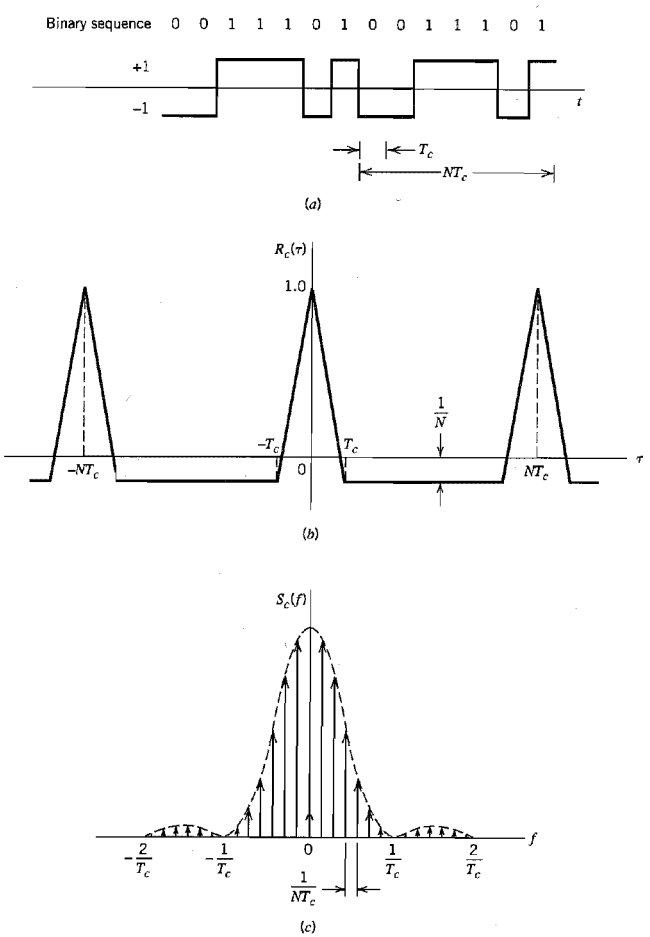


FIGURE 7.3 (*a*) Waveform of maximal-length sequence for length $m = 3$ or period $N = 7$. (*b*) Autocorrelation function. (*c*) Power spectral density. All three parts refer to the output of the feedback shift register of Figure 7.2.

where the lag τ lies in the interval $(-T_b/2, T_b/2)$; Equation (7.4) is a special case of Equation (1.26). Applying this formula to a maximal-length sequence represented by $c(t)$, we get

$$R_c(\tau) = \begin{cases} 1 - \dfrac{N+1}{NT_c}|\tau|, & |\tau| \le T_c \\ -\dfrac{1}{N}, & \text{for the remainder of the period} \end{cases} \tag{7.5}$$

This result is plotted in Figure 7.3*b* for the case of $m = 3$ or $N = 7$.

From Fourier transform theory we know that periodicity in the time domain is transformed into uniform sampling in the frequency domain. This interplay between the time and frequency domains is borne out by the power spectral density of the maximal-length wave $c(t)$. Specifically, taking the Fourier transform of Equation (7.5), we get the sampled spectrum

$$S_c(f) = \frac{1}{N^2}\,\delta(f) + \frac{1+N}{N^2}\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \operatorname{sinc}^2\left(\frac{n}{N}\right)\delta\left(f - \frac{n}{NT_c}\right) \tag{7.6}$$

which is plotted in Figure 7.3c for $m = 3$ or $N = 7$.

Comparing the results of Figure 7.3 for a maximal-length sequence with the corresponding results shown in Figure 1.11 for a random binary sequence, we may make the following observations:

- For a period of the maximal-length sequence, the autocorrelation function $R_c(\tau)$ is somewhat similar to that of a random binary wave.
- The waveforms of both sequences have the same envelope, $\operatorname{sinc}^2(fT)$, for their power spectral densities. The fundamental difference between them is that whereas the random binary sequence has a continuous spectral density characteristic, the corresponding characteristic of a maximal-length sequence consists of delta functions spaced $1/NT_c$ Hz apart.

As the shift-register length m, or equivalently, the period N of the maximal-length sequence is increased, the maximal-length sequence becomes increasingly similar to the random binary sequence. Indeed, in the limit, the two sequences become identical when N is made infinitely large. However, the price paid for making N large is an increasing storage requirement, which imposes a practical limit on how large N can actually be made.

■ CHOOSING A MAXIMAL-LENGTH SEQUENCE

Now that we understand the properties of a maximal-length sequence and the fact that we can generate it using a linear feedback shift register, the key question that we need to address is: How do we find the feedback logic for a desired period N? The answer to this

TABLE 7.1 *Maximal-length sequences of shift-register lengths 2–8*

Shift-Register Length, m	*Feedback Taps*
2*	[2, 1]
3*	[3, 1]
4	[4, 1]
5*	[5, 2], [5, 4, 3, 2], [5, 4, 2, 1]
6	[6, 1], [6, 5, 2, 1], [6, 5, 3, 2]
7*	[7, 1], [7, 3], [7, 3, 2, 1], [7, 4, 3, 2], [7, 6, 4, 2], [7, 6, 3, 1], [7, 6, 5, 2], [7, 6, 5, 4, 2, 1], [7, 5, 4, 3, 2, 1]
8	[8, 4, 3, 2], [8, 6, 5, 3], [8, 6, 5, 2], [8, 5, 3, 1], [8, 6, 5, 1], [8, 7, 6, 1], [8, 7, 6, 5, 2, 1], [8, 6, 4, 3, 2, 1]

question is to be found in the theory of error-control codes, which is covered in Chapter 10. The task of finding the required feedback logic is made particularly easy for us by virtue of the extensive tables of the necessary feedback connections for varying shift-register lengths that have been compiled in the literature. In Table 7.1, we present the sets of maximal (feedback) taps pertaining to shift-register lengths $m = 2, 3, \ldots, 8$.[3] Note that as m increases, the number of alternative schemes (codes) is enlarged. Also, for every set of feedback connections shown in this table, there is an "image" set that generates an identical maximal-length code, reversed in time sequence.

The particular sets identified with an asterisk in Table 7.1 correspond to *Mersenne prime length sequences*, for which the period N is a prime number.

EXAMPLE 7.2

Consider a maximal-length sequence requiring the use of a linear feedback-shift register of length $m = 5$. For feedback taps, we select the set [5, 2] from Table 7.1. The corresponding configuration of the code generator is shown in Figure 7.4*a*. Assuming that the initial state is 10000, the evolution of one period of the maximal-length sequence generated by this scheme is shown in Table 7.2*a*, where we see that the generator returns to the initial 10000 after 31 iterations; that is, the period is 31, which agrees with the value obtained from Equation (7.2).

Suppose next we select another set of feedback taps from Table 7.1, namely, [5, 4, 2, 1]. The corresponding code generator is thus as shown in Figure 7.4*b*. For the initial state 10000, we now find that the evolution of the maximal-length sequence is as shown in Table 7.2*b*. Here again, the generator returns to the initial state 10000 after 31 iterations, and so it should. But the maximal-length sequence generated is different from that shown in Table 7.2*a*.

Clearly, the code generator of Figure 7.4*a* has an advantage over that of Figure 7.4*b*, as it requires fewer feedback connections. ◀

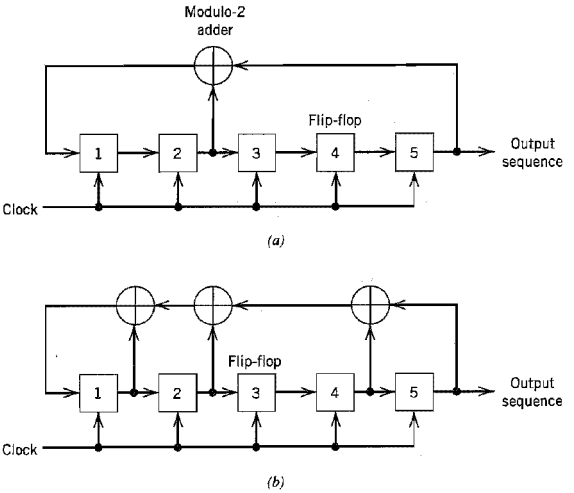


FIGURE 7.4 Two different configurations of feedback shift register of length $m = 5$. (*a*) Feedback connections [5, 2]. (*b*) Feedback connections [5, 4, 2, 1].

TABLE 7.2*a* ***Evolution of the maximal-length sequence generated by the feedback-shift register of Fig. 7.4a***

Feedback Symbol	*State of Shift Register*					*Output Symbol*
	1	0	0	0	0	
0	0	1	0	0	0	0
1	1	0	1	0	0	0
0	0	1	0	1	0	0
1	1	0	1	0	1	0
1	1	1	0	1	0	1
1	1	1	1	0	1	0
0	0	1	1	1	0	1
1	1	0	1	1	1	0
1	1	1	0	1	1	1
0	0	1	1	0	1	1
0	0	0	1	1	0	1
0	0	0	0	1	1	0
1	1	0	0	0	1	1
1	1	1	0	0	0	1
1	1	1	1	0	0	0
1	1	1	1	1	0	0
1	1	1	1	1	1	0
0	0	1	1	1	1	1
0	0	0	1	1	1	1
1	1	0	0	1	1	1
1	1	1	0	0	1	1
0	0	1	1	0	0	1
1	1	0	1	1	0	0
0	0	1	0	1	1	0
0	0	0	1	0	1	1
1	1	0	0	1	0	1
0	0	1	0	0	1	0
0	0	0	1	0	0	1
0	0	0	0	1	0	0
0	0	0	0	0	1	0
1	1	0	0	0	0	1

Code: 0000101011101100011111001101001

TABLE 7.2*b* ***Evolution of the maximal-length sequence generated by the feedback-shift register of Fig. 7.4b***

Feedback Symbol	*State of Shift Register*					*Output Symbol*
	1	0	0	0	0	
1	1	1	0	0	0	0
0	0	1	1	0	0	0
1	1	0	1	1	0	0
0	0	1	0	1	1	0
1	1	0	1	0	1	1
0	0	1	0	1	0	1
0	0	0	1	0	1	0
1	1	0	0	1	0	1
0	0	1	0	0	1	0
0	0	0	1	0	0	1
0	0	0	0	1	0	0
1	1	0	0	0	1	0
0	0	1	0	0	0	1
1	1	0	1	0	0	0
1	1	1	0	1	0	0
1	1	1	1	0	1	0
1	1	1	1	1	0	1
1	1	1	1	1	1	0
0	0	1	1	1	1	1
1	1	0	1	1	1	1
1	1	1	0	1	1	1
0	0	1	1	0	1	1
0	0	0	1	1	0	1
1	1	0	0	1	1	0
1	1	1	0	0	1	1
1	1	1	1	0	0	1
0	0	1	1	1	0	0
0	0	0	1	1	1	0
0	0	0	0	1	1	1
0	0	0	0	0	1	1
1	1	0	0	0	0	1

Code: 0000110101001000101111101100111

7.3 A Notion of Spread Spectrum

An important attribute of spread-spectrum modulation is that it can provide protection against externally generated interfering (jamming) signals with finite power. The jamming signal may consist of a fairly powerful broadband noise or multitone waveform that is directed at the receiver for the purpose of disrupting communications. Protection against jamming waveforms is provided by purposely making the information-bearing signal occupy a bandwidth far in excess of the minimum bandwidth necessary to transmit it. This has the effect of making the transmitted signal assume a noiselike appearance so as to blend into the background. The transmitted signal is thus enabled to propagate through the channel undetected by anyone who may be listening. We may therefore think of spread spectrum as a method of "camouflaging" the information-bearing signal.

One method of widening the bandwidth of an information-bearing (data) sequence involves the use of *modulation*. Let $\{b_k\}$ denote a binary data sequence, and $\{c_k\}$ denote a pseudo-noise (PN) sequence. Let the waveforms $b(t)$ and $c(t)$ denote their respective polar nonreturn-to-zero representations in terms of two levels equal in amplitude and opposite in polarity, namely, ± 1. We will refer to $b(t)$ as the information-bearing (data) signal, and to $c(t)$ as the PN signal. The desired modulation is achieved by applying the data signal $b(t)$ and the PN signal $c(t)$ to a product modulator or multiplier, as in Figure 7.5*a*. We know from Fourier transform theory that multiplication of two signals produces a signal whose spectrum equals the convolution of the spectra of the two component signals. Thus, if the message signal $b(t)$ is narrowband and the PN signal $c(t)$ is wideband, *the product (modulated) signal $m(t)$ will have a spectrum that is nearly the same as the wideband PN signal*. In other words, in the context of our present application, the PN sequence performs the role of a *spreading code*.

By multiplying the information-bearing signal $b(t)$ by the PN signal $c(t)$, each information bit is "chopped" up into a number of small time increments, as illustrated in the waveforms of Figure 7.6. These small time increments are commonly referred to as *chips*.

For *baseband* transmission, the product signal $m(t)$ represents the *transmitted signal*. We may thus express the transmitted signal as

$$m(t) = c(t)b(t) \tag{7.7}$$

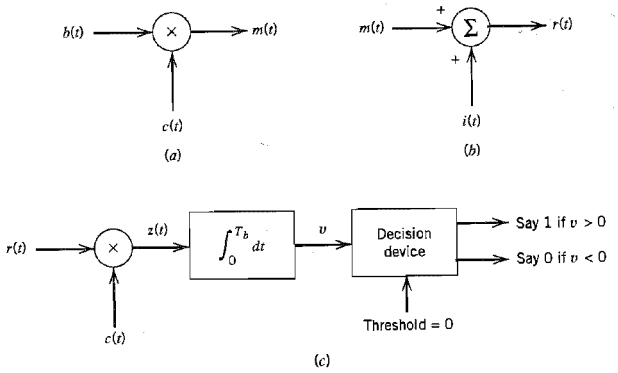
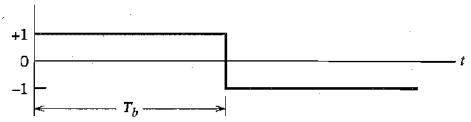
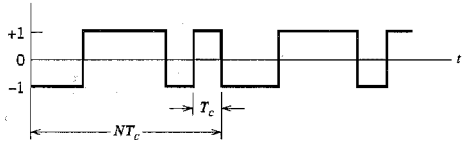


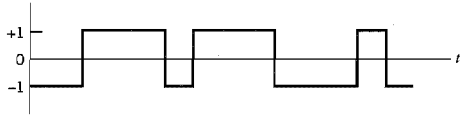
FIGURE 7.5 Idealized model of baseband spread-spectrum system. (*a*) Transmitter. (*b*) Channel. (*c*) Receiver.



(*a*) Data signal $b(t)$



(*b*) Spreading code $c(t)$



(*c*) Product signal $m(t)$

FIGURE 7.6 Illustrating the waveforms in the transmitter of Figure 7.5*a*.

The received signal $r(t)$ consists of the transmitted signal $m(t)$ plus an additive *interference* denoted by $i(t)$, as shown in the channel model of Figure 7.5*b*. Hence,

$$\begin{aligned} r(t) &= m(t) + i(t) \\ &= c(t)b(t) + i(t) \end{aligned} \tag{7.8}$$

To recover the original message signal $b(t)$, the received signal $r(t)$ is applied to a *demodulator* that consists of a multiplier followed by an integrator, and a decision device, as in Figure 7.5*c*. The multiplier is supplied with a locally generated PN sequence that is an exact *replica* of that used in the transmitter. Moreover, we assume that the receiver operates in perfect *synchronism* with the transmitter, which means that the PN sequence in the receiver is lined up exactly with that in the transmitter. The multiplier output in the receiver is therefore given by

$$\begin{aligned} z(t) &= c(t)r(t) \\ &= c^2(t)b(t) + c(t)i(t) \end{aligned} \tag{7.9}$$

Equation (7.9) shows that the data signal $b(t)$ is multiplied *twice* by the PN signal $c(t)$, whereas the unwanted signal $i(t)$ is multiplied only *once*. The PN signal $c(t)$ alternates between the levels -1 and $+1$, and the alternation is destroyed when it is squared; hence,

$$c^2(t) = 1 \qquad \text{for all } t \tag{7.10}$$

Accordingly, we may simplify Equation (7.9) as

$$z(t) = b(t) + c(t)i(t) \tag{7.11}$$

 WEEK 12

We thus see from Equation (7.11) that the data signal $b(t)$ is reproduced at the multiplier output in the receiver, except for the effect of the interference represented by the additive term $c(t)i(t)$. Multiplication of the interference $i(t)$ by the locally generated PN signal $c(t)$ means that the spreading code will affect the interference just as it did the original signal at the transmitter. We now observe that the data component $b(t)$ is narrowband, whereas the spurious component $c(t)i(t)$ is wideband. Hence, by applying the multiplier output to a baseband (low-pass) filter with a bandwidth just large enough to accommodate the recovery of the data signal $b(t)$, most of the power in the spurious component $c(t)i(t)$ is filtered out. The effect of the interference $i(t)$ is thus significantly reduced at the receiver output.

In the receiver shown in Figure 7.5*c*, the low-pass filtering action is actually performed by the integrator that evaluates the area under the signal produced at the multiplier output. The integration is carried out for the bit interval $0 \leq t \leq T_b$, providing the sample value v. Finally, a decision is made by the receiver: If v is greater than the threshold of zero, the receiver says that binary symbol 1 of the original data sequence was sent in the interval $0 \leq t \leq T_b$, and if v is less than zero, the receiver says that symbol 0 was sent; if v is exactly zero the receiver makes a random guess in favor of 1 or 0.

In summary, the use of a spreading code (with pseudo-random properties) in the transmitter produces a wideband transmitted signal that appears *noiselike* to a receiver that has *no* knowledge of the spreading code. From the discussion presented in Section 7.2, we recall that (for a prescribed data rate) the longer we make the period of the spreading code, the closer will the transmitted signal be to a truly random binary wave, and the harder it is to detect. Naturally, the price we have to pay for the improved protection against interference is increased transmission bandwidth, system complexity, and processing delay. However, when our primary concern is the security of transmission, these are not unreasonable costs to pay.

7.4 Direct-Sequence Spread Spectrum with Coherent Binary Phase-Shift Keying

The spread-spectrum technique described in the previous section is referred to as *direct-sequence spread spectrum*. The discussion presented there was in the context of baseband transmission. To provide for the use of this technique in passband transmission over a satellite channel, for example, we may incorporate *coherent binary phase-shift keying* (PSK) into the transmitter and receiver, as shown in Figure 7.7. The transmitter of Figure 7.7*a* first converts the incoming binary data sequence $\{b_k\}$ into a polar NRZ waveform $b(t)$, which is followed by two stages of modulation. The first stage consists of a product modulator or multiplier with the data signal $b(t)$ (representing a data sequence) and the PN signal $c(t)$ (representing the PN sequence) as inputs. The second stage consists of a binary PSK modulator. The transmitted signal $x(t)$ is thus a *direct-sequence spread binary phase-shift-keyed* (DS/BPSK) *signal*. The phase modulation $\theta(t)$ of $x(t)$ has one of two values, 0 and π, depending on the polarities of the message signal $b(t)$ and PN signal $c(t)$ at time t in accordance with the truth table of Table 7.3.

Figure 7.8 illustrates the waveforms for the second stage of modulation. Part of the modulated waveform shown in Figure 7.6*c* is reproduced in Figure 7.8*a*; the waveform shown here corresponds to one period of the PN sequence. Figure 7.8*b* shows the waveform of a sinusoidal carrier, and Figure 7.8*c* shows the DS/BPSK waveform that results from the second stage of modulation.

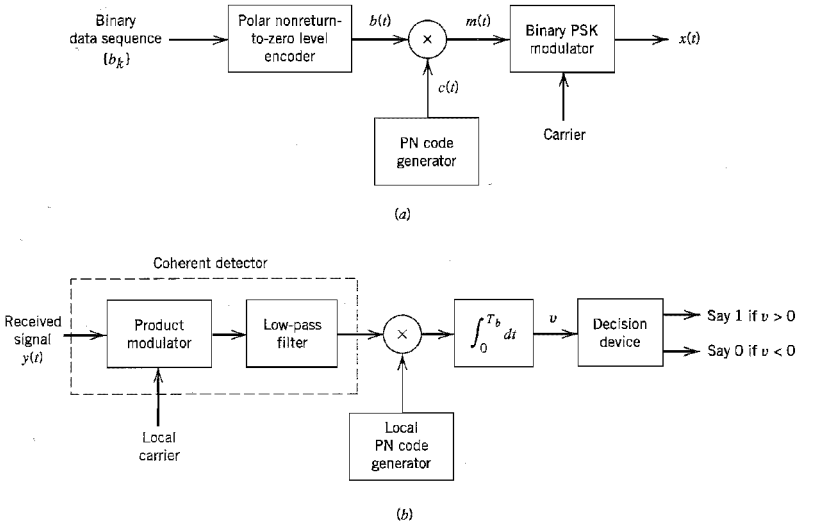


FIGURE 7.7 Direct-sequence spread coherent phase-shift keying. (*a*) Transmitter. (*b*) Receiver.

The receiver, shown in Figure 7.7*b*, consists of two stages of demodulation. In the first stage, the received signal $y(t)$ and a locally generated carrier are applied to a product modulator followed by a low-pass filter whose bandwidth is equal to that of the original message signal $m(t)$. This stage of the demodulation process reverses the phase-shift keying applied to the transmitted signal. The second stage of demodulation performs spectrum despreading by multiplying the low-pass filter output by a locally generated replica of the PN signal $c(t)$, followed by integration over a bit interval $0 \leq t \leq T_b$, and finally decision-making in the manner described in Section 7.3.

Model for Analysis

In the normal form of the transmitter, shown in Figure 7.7*a*, the spectrum spreading is performed prior to phase modulation. For the purpose of analysis, however, we find it more convenient to interchange the order of these operations, as shown in the model of

TABLE 7.3 *Truth table for phase modulation* $\theta(t)$, *radians*

		Polarity of Data Sequence b(t) *at Time* t	
		+	−
Polarity of PN sequence c(t) *at time* t	+	0	π
	−	π	0

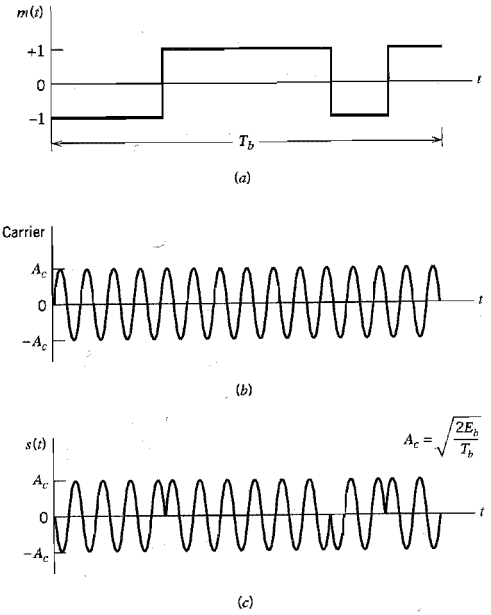


FIGURE 7.8 (*a*) Product signal $m(t) = c(t)b(t)$. (*b*) Sinusoidal carrier. (*c*) DS/BPSK signal.

Figure 7.9. We are permitted to do this because the spectrum spreading and the binary phase-shift keying are both linear operations; likewise for the phase demodulation and spectrum despreading. But for the interchange of operations to be feasible, it is important to synchronize the incoming data sequence and the PN sequence. The model of Figure 7.9 also includes representations of the channel and the receiver. In this model, it is assumed that the interference $j(t)$ limits performance, so that the effect of channel noise may be ignored. Accordingly, the channel output is given by

$$\begin{aligned} y(t) &= x(t) + j(t) \\ &= c(t)s(t) + j(t) \end{aligned} \tag{7.12}$$

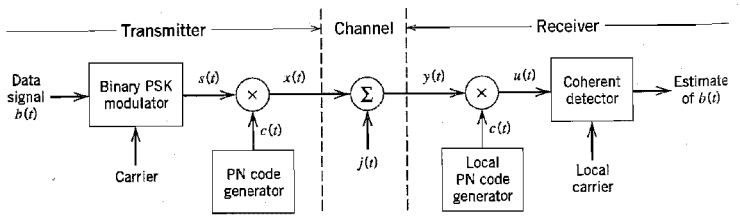


FIGURE 7.9 Model of direct-sequence spread binary PSK system.

where $s(t)$ is the binary PSK signal, and $c(t)$ is the PN signal. In the channel model included in Figure 7.9, the interfering signal is denoted by $j(t)$. This notation is chosen purposely to be different from that used for the interference in Figure 7.5*b*. The channel model in Figure 7.9 is passband in spectral content, whereas that in Figure 7.5*b* is in baseband form.

In the receiver, the received signal $y(t)$ is first multiplied by the PN signal $c(t)$ yielding an output that equals the coherent detector input $u(t)$. Thus,

$$\begin{aligned} u(t) &= c(t)y(t) \\ &= c^2(t)s(t) + c(t)j(t) \\ &= s(t) + c(t)j(t) \end{aligned} \tag{7.13}$$

In the last line of Equation (7.13), we have noted that, by design, the PN signal $c(t)$ satisfies the property described in Equation (7.10), reproduced here for convenience:

$$c^2(t) = 1 \qquad \text{for all } t$$

Equation (7.13) shows that the coherent detector input $u(t)$ consists of a binary PSK signal $s(t)$ embedded in additive code-modulated interference denoted by $c(t)j(t)$. The modulated nature of the latter component forces the interference signal (jammer) to spread its spectrum such that the detection of information bits at the receiver output is afforded increased reliability.

■ SYNCHRONIZATION

For its proper operation, a spread-spectrum communication system requires that the locally generated PN sequence used in the receiver to despread the received signal be *synchronized* to the PN sequence used to spread the transmitted signal in the transmitter.[4] A solution to the synchronization problem consists of two parts: *acquisition* and *tracking*. In acquisition, or *coarse* synchronization, the two PN codes are aligned to within a fraction of the chip in as short a time as possible. Once the incoming PN code has been acquired, tracking, or *fine* synchronization, takes place. Typically, PN acquisition proceeds in two steps. First, the received signal is multiplied by a locally generated PN code to produce a measure of *correlation* between it and the PN code used in the transmitter. Next, an appropriate *decision-rule and search strategy* is used to process the measure of correlation so obtained to determine whether the two codes are in synchronism and what to do if they are not. As for tracking, it is accomplished using phase-lock techniques very similar to those used for the local generation of coherent carrier references. The principal difference between them lies in the way in which phase discrimination is implemented.

7.5 *Signal-Space Dimensionality and Processing Gain*

Having developed a conceptual understanding of spread-spectrum modulation and a method for its implementation, we are ready to undertake a detailed mathematical analysis of the technique. The approach we have in mind is based on the signal-space theoretic ideas of Chapter 5. In particular, we develop signal-space representations of the transmitted signal and the interfering signal (jammer).

In this context, consider the set of orthonormal basis functions:

$$\phi_k(t) = \begin{cases} \sqrt{\frac{2}{T_c}} \cos(2\pi f_c t), & kT_c \le t \le (k+1)T_c \\ 0, & \text{otherwise} \end{cases} \tag{7.14}$$

$$\tilde{\phi}_k(t) = \begin{cases} \sqrt{\frac{2}{T_c}} \sin(2\pi f_c t), & kT_c \le t \le (k+1)T_c \\ 0, & \text{otherwise} \end{cases} \tag{7.15}$$

$$k = 0, 1, \ldots, N-1$$

where T_c is the *chip duration*, and N is the number of chips per bit. Accordingly, we may describe the transmitted signal $x(t)$ for the interval of an information bit as follows:

$$\begin{aligned} x(t) &= c(t)s(t) \\ &= \pm\sqrt{\frac{2E_b}{T_b}}\, c(t) \cos(2\pi f_c t) \\ &= \pm\sqrt{\frac{E_b}{N}} \sum_{k=0}^{N-1} c_k \phi_k(t), \qquad 0 \le t \le T_b \end{aligned} \tag{7.16}$$

where E_b is the signal energy per bit; the plus sign corresponds to information bit 1, and the minus sign corresponds to information bit 0. The code sequence $\{c_0, c_1, \ldots, c_{N-1}\}$ denotes the PN sequence, with $c_k = \pm 1$. The transmitted signal $x(t)$ is therefore N-dimensional in that it requires a minimum of N orthonormal functions for its representation.

Consider next the representation of the interfering signal (jammer), $j(t)$. Ideally, the jammer likes to place all of its available energy in exactly the same N-dimensional signal space as the transmitted signal $x(t)$; otherwise, part of its energy goes to waste. However, the best that the jammer can hope to know is the transmitted signal bandwidth. Moreover, there is no way that the jammer can have knowledge of the signal phase. Accordingly, we may represent the jammer by the general form

$$j(t) = \sum_{k=0}^{N-1} j_k \phi_k(t) + \sum_{k=0}^{N-1} \tilde{j}_k \tilde{\phi}_k(t), \qquad 0 \le t \le T_b \tag{7.17}$$

where

$$j_k = \int_0^{T_b} j(t)\phi_k(t)\, dt, \qquad k = 0, 1, \ldots, N-1 \tag{7.18}$$

and

$$\tilde{j}_k = \int_0^{T_b} j(t)\tilde{\phi}_k(t)\, dt, \qquad k = 0, 1, \ldots, N-1 \tag{7.19}$$

Thus the interference $j(t)$ is $2N$-dimensional; that is, it has twice the number of dimensions required for representing the transmitted DS/BPSK signal $x(t)$. In terms of the represen-

tation given in Equation (7.17), we may express the average power of the interference $j(t)$ as follows:

$$\begin{aligned} J &= \frac{1}{T_b}\int_0^{T_b} j^2(t)\,dt \\ &= \frac{1}{T_b}\sum_{k=0}^{N-1} j_k^2 + \frac{1}{T_b}\sum_{k=0}^{N-1} \tilde{j}_k^2 \end{aligned} \tag{7.20}$$

Moreover, due to lack of knowledge of signal phase, the best strategy a jammer can apply is to place equal energy in the cosine and sine coordinates defined in Equations (7.18) and (7.19); hence, we may safely assume

$$\sum_{k=0}^{N-1} j_k^2 = \sum_{k=0}^{N-1} \tilde{j}_k^2 \tag{7.21}$$

Correspondingly, we may simplify Equation (7.20) as

$$J = \frac{2}{T_b}\sum_{k=0}^{N-1} j_k^2 \tag{7.22}$$

Our aim is to tie these results together by finding the signal-to-noise ratios measured at the input and output of the DS/BPSK receiver in Figure 7.9. To that end, we use Equation (7.13) to express the coherent detector output as

$$\begin{aligned} v &= \sqrt{\frac{2}{T_b}}\int_0^{T_b} u(t)\cos(2\pi f_c t)\,dt \\ &= v_s + v_{cj} \end{aligned} \tag{7.23}$$

where the components v_s and v_{cj} are due to the despread binary PSK signal, $s(t)$, and the spread interference, $c(t)j(t)$, respectively. These two components are defined as follows:

$$v_s = \sqrt{\frac{2}{T_b}}\int_0^{T_b} s(t)\cos(2\pi f_c t)\,dt \tag{7.24}$$

and

$$v_{cj} = \sqrt{\frac{2}{T_b}}\int_0^{T_b} c(t)j(t)\cos(2\pi f_c t)\,dt \tag{7.25}$$

Consider first the component v_s due to the signal. The despread binary PSK signal $s(t)$ equals

$$s(t) = \pm\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t), \qquad 0 \le t \le T_b \tag{7.26}$$

where the plus sign corresponds to information bit 1, and the minus sign corresponds to information bit 0. Hence, assuming that the carrier frequency f_c is an integer multiple of $1/T_b$, we have

$$v_s = \pm\sqrt{E_b} \tag{7.27}$$

Consider next the component v_{cj} due to interference. Expressing the PN signal $c(t)$ in the explicit form of a sequence, $\{c_0, c_1, \ldots, c_{N-1}\}$, we may rewrite Equation (7.25) in the corresponding form

$$v_{cj} = \sqrt{\frac{2}{T_b}} \sum_{k=0}^{N-1} c_k \int_{kT_c}^{(k+1)T_c} j(t) \cos(2\pi f_c t)\, dt \tag{7.28}$$

Using Equation (7.14) for $\phi_k(t)$, and then Equation (7.18) for the coefficient j_k, we may redefine v_{cj} as

$$\begin{aligned} v_{cj} &= \sqrt{\frac{T_c}{T_b}} \sum_{k=0}^{N-1} c_k \int_0^{T_b} j(t)\phi_k(t)\, dt \\ &= \sqrt{\frac{T_c}{T_b}} \sum_{k=0}^{N-1} c_k j_k \end{aligned} \tag{7.29}$$

We next approximate the PN sequence as an *independent and identically distributed (i.i.d.) binary sequence*. We emphasize the implication of this approximation by recasting Equation (7.29) in the form

$$V_{cj} = \sqrt{\frac{T_c}{T_b}} \sum_{k=0}^{N-1} C_k j_k \tag{7.30}$$

where V_{cj} and C_k are random variables with sample values v_{cj} and c_k, respectively. In Equation (7.30), the jammer is assumed to be fixed. With the C_k treated as i.i.d. random variables, we find that the probability of the event $C_k = \pm 1$ is

$$P(C_k = 1) = P(C_k = -1) = \tfrac{1}{2} \tag{7.31}$$

Accordingly, the mean of the random variable V_{cj} is zero since, for fixed k, we have

$$\begin{aligned} E[C_k j_k | j_k] &= j_k P(C_k = 1) - j_k P(C_k = -1) \\ &= \tfrac{1}{2} j_k - \tfrac{1}{2} j_k \\ &= 0 \end{aligned} \tag{7.32}$$

For a fixed vector $\mathbf{j}$, representing the set of coefficients $j_0, j_1, \ldots, j_{N-1}$, the variance of V_{cj} is given by

$$\operatorname{var}[V_{cj} | \mathbf{j}] = \frac{1}{N} \sum_{k=0}^{N-1} j_k^2 \tag{7.33}$$

Since the *spread factor* $N = T_b/T_c$, we may use Equation (7.22) to express this variance in terms of the average interference power J as

$$\operatorname{var}[V_{cj} | \mathbf{j}] = \frac{JT_c}{2} \tag{7.34}$$

Thus the random variable V_{cj} has zero mean and variance $JT_c/2$.

From Equation (7.27), we note that the signal component at the coherent detector output (during each bit interval) equals $\pm\sqrt{E_b}$, where E_b is the signal energy per bit. Hence, the peak instantaneous power of the signal component is E_b. Accordingly, we may define

the *output signal-to-noise ratio* as the instantaneous peak power E_b divided by the variance of the equivalent noise component in Equation (7.34). We thus write

$$(\text{SNR})_O = \frac{2E_b}{JT_c} \tag{7.35}$$

The average signal power at the receiver input equals E_b/T_b. We thus define an *input signal-to-noise* ratio as

$$(\text{SNR})_I = \frac{E_b/T_b}{J} \tag{7.36}$$

Hence, eliminating E_b/J between Equations (7.35) and (7.36), we may express the output signal-to-noise ratio in terms of the input signal-to-noise ratio as

$$(\text{SNR})_O = \frac{2T_b}{T_c}(\text{SNR})_I \tag{7.37}$$

It is customary practice to express signal-to-noise ratios in decibels. To that end, we introduce a term called the *processing gain* (PG), which is defined as *the gain in SNR obtained by the use of spread spectrum.* Specifically, we write

$$\text{PG} = \frac{T_b}{T_c} \tag{7.38}$$

which represents the gain achieved by processing a spread-spectrum signal over an unspread signal. We may thus write Equation (7.37) in the equivalent form:

$$10 \log_{10}(\text{SNR})_O = 10 \log_{10}(\text{SNR})_I + 3 + 10 \log_{10}(\text{PG}) \text{ dB} \tag{7.39}$$

The 3-dB term on the right-hand side of Equation (7.39) accounts for the gain in SNR that is obtained through the use of coherent detection (which presumes exact knowledge of the signal phase by the receiver). This gain in SNR has nothing to do with the use of spread spectrum. Rather, it is the last term, $10 \log_{10}(\text{PG})$, that accounts for the processing gain. Note that both the processing gain PG and the spread factor N (i.e., PN sequence length) equal the ratio T_b/T_c. Thus, the longer we make the PN sequence (or, correspondingly, the smaller the chip time T_c is), the larger will the processing gain be.

7.6 *Probability of Error*

Let the coherent detector output v in the direct-sequence spread BPSK system of Figure 7.9 denote the sample value of a random variable V. Let the equivalent noise component v_{cj} produced by external interference denote the sample value of a random variable V_{cj}. Then, from Equations (7.23) and (7.27) we deduce that

$$V = \pm\sqrt{E_b} + V_{cj} \tag{7.40}$$

where E_b is the transmitted signal energy per bit. The plus sign refers to sending symbol (information bit) 1, and the minus sign refers to sending symbol 0. The decision rule used by the coherent detector of Figure 7.9 is to declare that the received bit in an interval $(0, T_b)$ is 1 if the detector output exceeds a threshold of zero, and that it is 0 if the detector output is less than the threshold; if the detector output is exactly zero, the receiver makes a random guess in favor of 1 or 0. With both information bits assumed equally likely, we

find that (because of the symmetric nature of the problem) the average probability of error P_e is the same as the conditional probability of (say) the receiver making a decision in favor of symbol 1, given that symbol 0 was sent. That is,

$$\begin{aligned} P_e &= P(V > 0 | \text{symbol 0 was sent}) \\ &= P(V_{cj} > \sqrt{E_b}) \end{aligned} \tag{7.41}$$

Naturally, the probability of error P_e depends on the random variable V_{cj} defined by Equation (7.30). According to this definition, V_{cj} is the sum of N identically distributed random variables. Hence, from the *central limit theorem*, we deduce that for large N, the random variable V_{cj} assumes a Gaussian distribution. Indeed, the spread factor or PN sequence length N is typically large in the direct-sequence spread-spectrum systems encountered in practice, under which condition the application of the central limit theorem is justified.

Earlier we evaluated the mean and variance of V_{cj}; see Equations (7.32) and (7.34). We may therefore state that the equivalent noise component V_{cj} contained in the coherent detector output may be approximated as a Gaussian random variable with zero mean and variance $JT_c/2$, where J is the average interference power and T_c is the chip duration. With this approximation at hand, we may then proceed to calculate the probability of the event $V_{cj} > \sqrt{E_b}$, and thus express the average probability of error in accordance with Equation (7.41) as

$$P_e \simeq \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_b}{JT_c}}\right) \tag{7.42}$$

This simple formula, which invokes the Gaussian assumption, is appropriate for DS/BPSK binary systems with large spread factor N.

■ ANTIJAM CHARACTERISTICS

It is informative to compare Equation (7.42) with the formula for the average probability of error for a coherent binary PSK system reproduced here for convenience of presentation [see Equation (6.20)]

$$P_e = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{7.43}$$

Based on this comparison, we see that insofar as the calculation of bit error rate in a direct-sequence spread binary PSK system is concerned, the interference may be treated as wideband noise of power spectral density $N_0/2$, defined by

$$\frac{N_0}{2} = \frac{JT_c}{2} \tag{7.44}$$

This relation is simply a restatement of an earlier result given in Equation (7.34).

Since the signal energy per bit $E_b = PT_b$, where P is the average signal power and T_b is the bit duration, we may express the signal energy per bit-to-noise spectral density ratio as

$$\frac{E_b}{N_0} = \left(\frac{T_b}{T_c}\right)\left(\frac{P}{J}\right) \tag{7.45}$$

Using the definition of Equation (7.38) for the processing gain PG we may reformulate this result as

$$\frac{J}{P} = \frac{\text{PG}}{E_b/N_0} \tag{7.46}$$

The ratio J/P is termed the *jamming margin*. Accordingly, the jamming margin and the processing gain, both expressed in decibels, are related by

$$(\text{Jamming margin})_{\text{dB}} = (\text{Procesing gain})_{\text{dB}} - 10 \log_{10}\left(\frac{E_b}{N_0}\right)_{\text{min}} \tag{7.47}$$

where $(E_b/N_0)_{\text{min}}$ is the minimum value needed to support a prescribed average probability of error.

EXAMPLE 7.3

A spread-spectrum communication system has the following parameters:

$$\text{Information bit duration, } T_b = 4.095 \text{ ms}$$
$$\text{PN chip duration, } T_c = 1\ \mu\text{s}$$

Hence, using Equation (7.38) we find that the processing gain is

$$\text{PG} = 4095$$

Correspondingly, the required period of the PN sequence is $N = 4095$, and the shift-register length is $m = 12$.

For a satisfactory reception, we may assume that the average probability of error is not to exceed 10^{-5}. From the formula for a coherent binary PSK receiver, we find that $E_b/N_0 = 10$ yields an average probability of error equal to 0.387×10^{-5}. Hence, using this value for E_b/N_0, and the value calculated for the processing gain, we find from Equation (7.47) that the jamming margin is

$$\begin{aligned}(\text{Jamming margin})_{\text{dB}} &= 10 \log_{10} 4095 - 10 \log_{10}(10) \\ &= 36.1 - 10 \\ &= 26.1 \text{ dB}\end{aligned}$$

That is, information bits at the receiver output can be detected reliably even when the noise or interference at the receiver input is up to 409.5 times the received signal power. Clearly, this is a powerful advantage against interference (jamming), which is realized through the clever use of spread-spectrum modulation. ◀

7.7 *Frequency-Hop Spread Spectrum*

In the type of spread-spectrum systems discussed in Section 7.4, the use of a PN sequence to modulate a phase-shift-keyed signal achieves *instantaneous* spreading of the transmission bandwidth. The ability of such a system to combat the effects of jammers is determined by the processing gain of the system, which is a function of the PN sequence period. The processing gain can be made larger by employing a PN sequence with narrow chip duration, which, in turn, permits a greater transmission bandwidth and more chips per bit. However, the capabilities of physical devices used to generate the PN spread-spectrum signals impose a practical limit on the attainable processing gain. Indeed, it may turn out that the processing gain so attained is still not large enough to overcome the effects of

some jammers of concern, in which case we have to resort to other methods. One such alternative method is to force the jammer to cover a wider spectrum by *randomly hopping* the data-modulated carrier from one frequency to the next. In effect, the spectrum of the transmitted signal is spread *sequentially* rather than instantaneously; the term "sequentially" refers to the pseudo-random-ordered sequence of frequency hops.

The type of spread spectrum in which the carrier hops randomly from one frequency to another is called *frequency-hop (FH) spread spectrum*. A common modulation format for FH systems is that of *M-ary frequency-shift keying* (MFSK). The combination of these two techniques is referred to simply as FH/MFSK. (A description of M-ary FSK is presented in Chapter 6.)

Since frequency hopping does not cover the entire spread spectrum instantaneously, we are led to consider the rate at which the hops occur. In this context, we may identify two basic (technology-independent) characterizations of frequency hopping:

1. *Slow-frequency hopping*, in which the *symbol rate* R_s of the MFSK signal is an integer multiple of the *hop rate* R_h. That is, several symbols are transmitted on each frequency hop.
2. *Fast-frequency hopping*, in which the hop rate R_h is an integer multiple of the MFSK symbol rate R_s. That is, the carrier frequency will change or hop several times during the transmission of one symbol.

Obviously, slow-frequency hopping and fast-frequency hopping are the converse of one another. In the following, these two characterizations of frequency hopping are considered in turn.

■ SLOW-FREQUENCY HOPPING

Figure 7.10*a* shows the block diagram of an FH/MFSK transmitter, which involves *frequency modulation* followed by *mixing*. First, the incoming binary data are applied to an M-ary FSK modulator. The resulting modulated wave and the output from a digital *frequency synthesizer* are then applied to a mixer that consists of a multiplier followed by a band-pass filter. The filter is designed to select the sum frequency component resulting from the multiplication process as the transmitted signal. In particular, successive k-bit segments of a PN sequence drive the frequency synthesizer, which enables the carrier frequency to hop over 2^k distinct values. On a single hop, the bandwidth of the transmitted signal is the same as that resulting from the use of a conventional MFSK with an alphabet of $M = 2^K$ orthogonal signals. However, for a complete range of 2^k frequency hops, the transmitted FH/MFSK signal occupies a much larger bandwidth. Indeed, with present-day technology, FH bandwidths on the order of several GHz are attainable, which is an order of magnitude larger than that achievable with direct-sequence spread spectra. An implication of these large FH bandwidths is that coherent detection is possible only within each hop, because frequency synthesizers are unable to maintain phase coherence over successive hops. Accordingly, most frequency-hop spread-spectrum communication systems use noncoherent M-ary modulation schemes.

In the receiver depicted in Figure 7.10*b*, the frequency hopping is first removed by *mixing* (down-converting) the received signal with the output of a local frequency synthesizer that is synchronously controlled in the same manner as that in the transmitter. The resulting output is then band-pass filtered, and subsequently processed by a *noncoherent* M-ary FSK detector. To implement this M-ary detector, we may use a bank of M noncoherent matched filters, each of which is matched to one of the MFSK tones. (Noncoherent matched filters are described in Chapter 6.) An estimate of the original symbol transmitted is obtained by selecting the largest filter output.

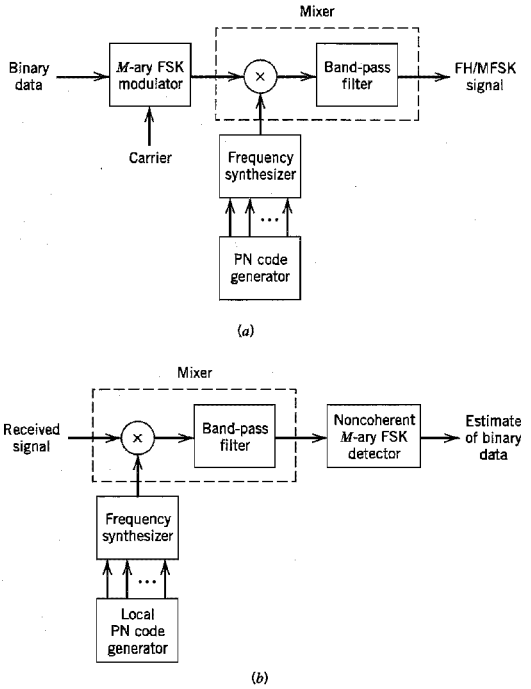


FIGURE 7.10 Frequency-hop spread M-ary frequency-shift keying. (*a*) Transmitter. (*b*) Receiver.

An individual FH/MFSK tone of shortest duration is referred to as a *chip*; this terminology should not be confused with that used in Section 7.4 describing DS/BPSK. The *chip rate*, R_c, for an FH/MFSK system is defined by

$$R_c = \max(R_h, R_s) \tag{7.48}$$

where R_h is the *hop rate*, and R_s is the *symbol rate*.

A slow FH/MFSK signal is characterized by having multiple symbols transmitted per hop. Hence, each symbol of a slow FH/MFSK signal is a chip. Correspondingly, in a slow FH/MFSK system, the bit rate R_b of the incoming binary data, the symbol rate R_s of the MFSK signal, the chip rate R_c, and the hop rate R_h are related by

$$R_c = R_s = \frac{R_b}{K} \geq R_h \tag{7.49}$$

where $K = \log_2 M$.

At each hop, the MFSK tones are separated in frequency by an integer multiple of the chip rate $R_c = R_s$, ensuring their orthogonality. The implication of this condition is that any transmitted symbol will not produce any crosstalk in the other $M - 1$ noncoherent matched filters constituting the MFSK detector of the receiver in Figure 7.10*b*. By "crosstalk" we mean the spillover from one filter output into an adjacent one. The resulting performance of the slow FH/MFSK system is the same as that for the noncoherent detection

of conventional (unhopped) MFSK signals in additive white Gaussian noise. Thus the interfering (jamming) signal has an effect on the FH/MFSK receiver, in terms of average probability of symbol error, equivalent to that of additive white Gaussian noise on a conventional noncoherent M-ary FSK receiver experiencing no interference. On the basis of this equivalence, we may use Equation (6.140) for approximate evaluation of the probability of symbol error in the FH/MFSK system.

Assuming that the jammer decides to spread its average power J over the entire frequency-hopped spectrum, the jammer's effect is equivalent to an AWGN with power spectral density $N_0/2$, where $N_0 = J/W_c$ and W_c is the FH bandwidth. The spread-spectrum system is thus characterized by the *symbol energy-to-noise spectral density ratio*:

$$\frac{E}{N_0} = \frac{P/J}{W_c/R_s} \tag{7.50}$$

where the ratio P/J is the reciprocal of the jamming margin. The other ratio in the denominator of Equation (7.50) is the processing gain of the slow FH/MFSK system, which is defined by

$$\begin{aligned} \text{PG} &= \frac{W_c}{R_s} \\ &= 2^k \end{aligned} \tag{7.51}$$

That is, the processing gain (expressed in decibels) is equal to $10 \log_{10} 2^k \simeq 3k$, where k is the length of the PN segment employed to select a frequency hop.

This result assumes that the jammer spreads its power over the entire FH spectrum. However, if the jammer decides to concentrate on just a few of the hopped frequencies, then the processing gain realized by the receiver would be less than $3k$ decibels.

▶ EXAMPLE 7.4

Figure 7.11*a* illustrates the variation of the frequency of a slow FH/MFSK signal with time for one complete period of the PN sequence. The period of the PN sequence is $2^4 - 1 = 15$. The FH/MFSK signal has the following parameters:

Number of bits per MFSK symbol	$K = 2$
Number of MFSK tones	$M = 2^K = 4$
Length of PN segment per hop	$k = 3$
Total number of frequency hops	$2^k = 8$

In this example, the carrier is hopped to a new frequency after transmitting two symbols or equivalently, four information bits. Figure 7.11*a* also includes the input binary data, and the PN sequence controlling the selection of FH carrier frequency. It is noteworthy that although there are eight distinct frequencies available for hopping, only three of them are utilized by the PN sequence.

Figure 7.11*b* shows the variation of the dehopped frequency with time. This variation is recognized to be the same as that of a conventional MFSK signal produced by the given input data. ◀

■ FAST-FREQUENCY HOPPING

A fast FH/MFSK system differs from a slow FH/MFSK system in that there are multiple hops per M-ary symbol. Hence, in a fast FH/MFSK system, each hop is a chip. In general,

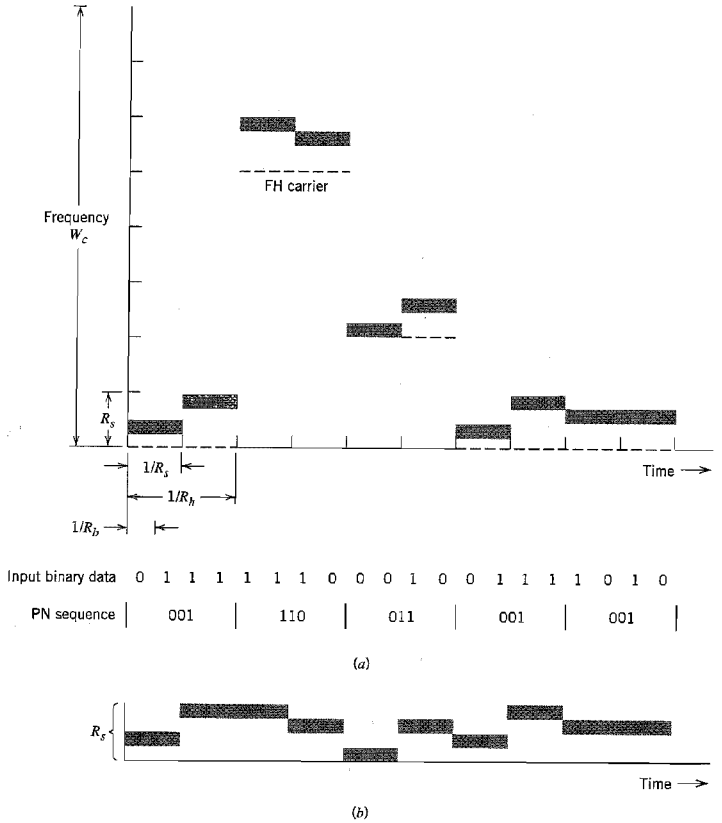


FIGURE 7.11 Illustrating slow-frequency hopping. (*a*) Frequency variation for one complete period of the PN sequence. (*b*) Variation of the dehopped frequency with time.

fast-frequency hopping is used to defeat a smart jammer's tactic that involves two functions: measurements of the spectral content of the transmitted signal, and retuning of the interfering signal to that portion of the frequency band. Clearly, to overcome the jammer, the transmitted signal must be hopped to a new carrier frequency *before* the jammer is able to complete the processing of these two functions.

For data recovery at the receiver, noncoherent detection is used. However, the detection procedure is quite different from that used in a slow FH/MFSK receiver. In particular, two procedures may be considered:

1. For each FH/MFSK symbol, separate decisions are made on the K frequency-hop chips received, and a simple rule based on *majority vote* is used to make an estimate of the dehopped MFSK symbol.
2. For each FH/MFSK symbol, likelihood functions are computed as functions of the total signal received over K chips, and the largest one is selected.

A receiver based on the second procedure is optimum in the sense that it minimizes the average probability of symbol error for a given E_b/N_0.

▶ EXAMPLE 7.5

Figure 7.12*a* illustrates the variation of the transmitted frequency of a fast FH/MFSK signal with time. The signal has the following parameters:

Number of bits per MFSK symbol	$K = 2$
Number of MFSK tones	$M = 2^K = 4$
Length of PN segment per hop	$k = 3$
Total number of frequency hops	$2^k = 8$

In this example, each MFSK symbol has the same number of bits and chips; that is, the chip rate R_c is the same as the bit rate R_b. After each chip, the carrier frequency of the transmitted MFSK signal is hopped to a different value, except for few occasions when the k-chip segment of the PN sequence repeats itself.

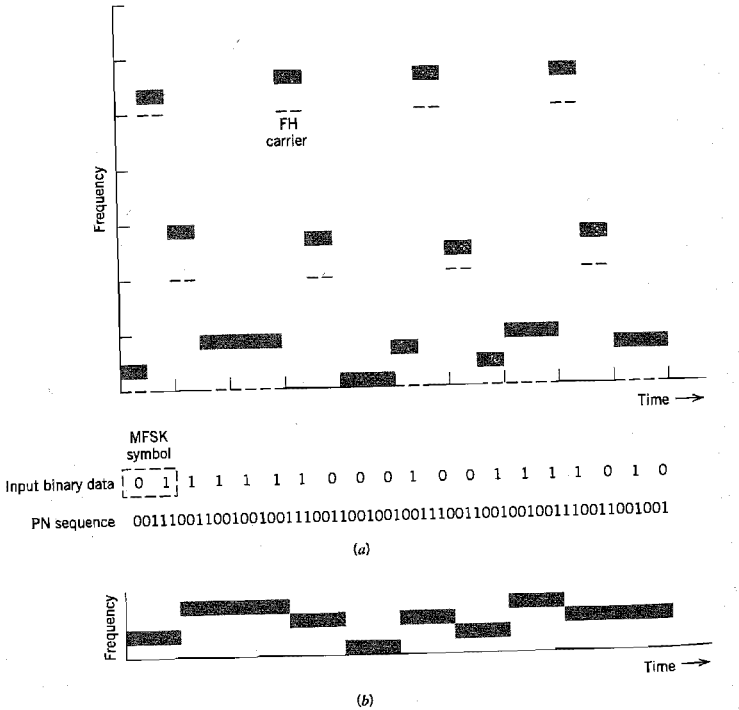


FIGURE 7.12 Illustrating fast-frequency hopping. (*a*) Variation of the transmitter frequency with time. (*b*) Variation of the dehopped frequency with time.

Figure 7.12*b* depicts the time variation of the frequency of the dehopped MFSK signal, which is the same as that in Example 7.4. ◀

7.8 *Computer Experiments: Maximal-Length and Gold Codes*

Code-division multiplexing (CDM) provides an alternative to the traditional methods of frequency-division multiplexing (FDM) and time-division multiplexing (TDM). It does not require the bandwidth allocation of FDM (discussed in Chapter 2) nor the time synchronization needed in TDM (discussed in Chapter 3). Rather, users of a common channel are permitted access to the channel through the assignment of a "spreading code" to each individual user under the umbrella of spread-spectrum modulation. The purpose of this computer experiment is to study a certain class of spreading codes for CDM systems that provide a satisfactory performance.

In an ideal CDM system, the cross-correlation between any two users of the system is zero. For this ideal condition to be realized, we require that the cross-correlation function between the spreading codes assigned to any two users of the system be zero for all cyclic shifts. Unfortunately, ordinary PN sequences do not satisfy this requirement because of their relatively poor cross-correlation properties.

As a remedy for this shortcoming of ordinary PN sequences, we may use a special class of PN sequences called *Gold sequences (codes)*,[5] the generation of which is embodied in the following theorem:

> Let $g_1(X)$ and $g_2(X)$ be a preferred pair of primitive polynomials of degree n whose corresponding shift registers generate maximal-length sequences of period $2^n - 1$ and whose cross-correlation function has a magnitude less than or equal to
>
> $$2^{(n+1)/2} + 1 \qquad \text{for } n \text{ odd} \tag{7.52}$$
>
> or
>
> $$2^{(n+2)/2} + 1 \qquad \text{for } n \text{ even and } n \neq 0 \bmod 4 \tag{7.53}$$
>
> Then the shift register corresponding to the product polynomial $g_1(X) \cdot g_2(X)$ will generate $2^n + 1$ different sequences, with each sequence having a period of $2^n - 1$, and the cross-correlation between any pair of such sequences satisfying the preceding condition.

Hereafter, this theorem is referred to as *Gold's theorem.*

To understand Gold's theorem, we need to define what we mean by a primitive polynomial. Consider a polynomial $g(X)$ defined over a *binary field* (i.e., a finite set of two elements, 0 and 1, which is governed by the rules of binary arithmetic). The polynomial $g(X)$ is said to be an *irreducible polynomial* if it cannot be factored using any polynomials from the binary field. An irreducible polynomial $g(X)$ of degree m is said to be a *primitive polynomial* if the smallest integer m for which the polynomial $g(X)$ divides the factor $X^n + 1$ is $n = 2^m - 1$. Further discussion of this topic is deferred to Chapter 8; in particular, see Example 8.3.

Experiment 1. Correlation Properties of PN Sequences

Consider a pair of shift registers for generating two PN sequences of period $2^7 - 1 = 127$. One feedback shift register has the feedback taps [7, 1] and the other one has the feedback taps [7, 6, 5, 4]. Both sequences have the same autocorrelation function shown in Figure 7.13*a*, which follows readily from the definition presented in Equation (7.5).

However, the calculation of the cross-correlation function between PN sequences is a more difficult proposition, particularly for large n. To perform this calculation, we resort to the use of computer simulation for varying cyclic shift τ inside the interval $0 < \tau \leq 2^n - 1$. The results of this computation are presented in Figure 7.13*b*. This figure confirms the poor cross-correlation property of PN sequences compared to their autocorrelation function. The magnitude of the cross-correlation function exceeds 40.

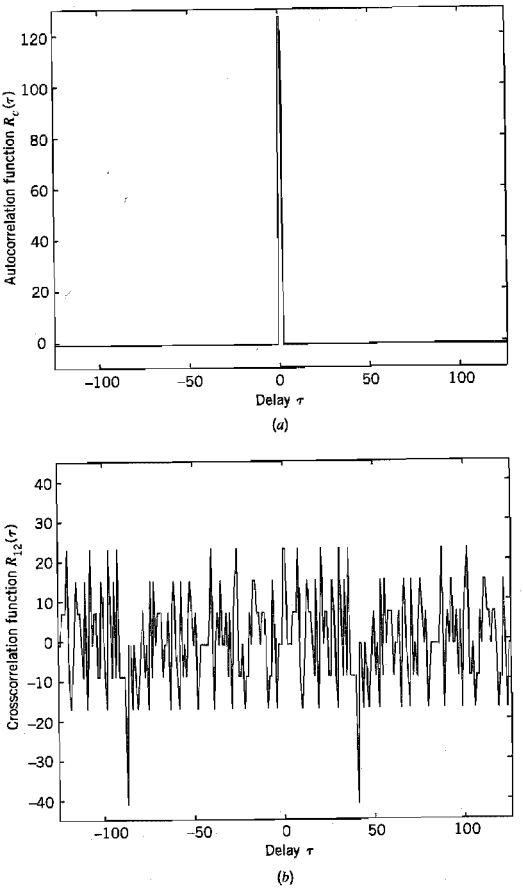


FIGURE 7.13 (*a*) Autocorrelation function $R_c(\tau)$, and (*b*) cross-correlation function $R_{12}(\tau)$ of the two PN sequences [7, 1] and [7, 6, 5, 4].

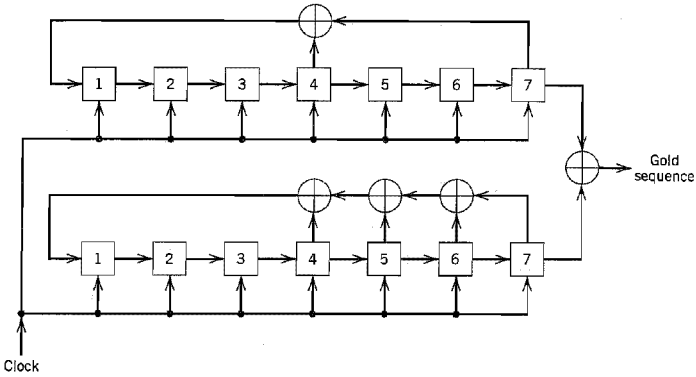


FIGURE 7.14 Generator for a Gold sequence of period $2^7 - 1 = 127$.

Experiment 2. Correlation Properties of Gold Sequences

For our next experiment, we consider Gold sequences with period $2^7 - 1 = 127$. To generate such a sequence for $n = 7$ we need a preferred pair of PN sequences that satisfy Equation (7.52) (n odd), as shown by

$$2^{(n+1)/2} + 1 = 2^4 + 1 = 17$$

This requirement is satisfied by the PN sequences with feedback taps [7, 4] and [7, 6, 5, 4]. The Gold-sequence generator is shown in Figure 7.14 that involves the modulo-2 addition of these two sequences. According to Gold's theorem, there are a total of

$$2^n + 1 = 2^7 + 1 = 129$$

sequences that satisfy Equation (7.52). The cross-correlation between any pair of such sequences is shown in Figure 7.15, which is indeed in full accord with Gold's theorem. In particular, the magnitude of the cross-correlation is less than or equal to 17.

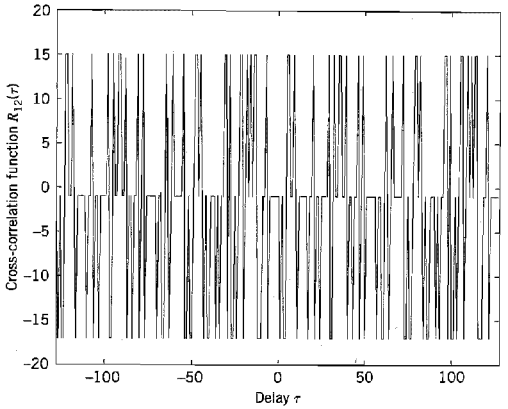


FIGURE 7.15 Cross-correlation function $R_{12}(\tau)$ of a pair of Gold sequences based on the two PN sequences [7, 4] and [7, 6, 5, 4].

7.9 Summary and Discussion

Direct-sequence M-ary phase shift keying (DS/MPSK) and *frequency-hop M-ary frequency shift-keying* (FH/MFSK) represent two principal categories of spread-spectrum communications. Both of them rely on the use of a pseudo-noise (PN) sequence, which is applied differently in the two categories.

In a DS/MPSK system, the PN sequence makes the transmitted signal assume a noiselike appearance by spreading its spectrum over a broad range of frequencies simultaneously. For the phase-shift keying, we may use binary PSK (i.e., $M = 2$) with a single carrier. Alternatively, we may use QPSK (i.e., $M = 4$), in which case the data are transmitted using a pair of carriers in phase quadrature. (Both PSK and QPSK are discussed in Section 6.3.) The usual motivation for using QPSK is to provide for improved bandwidth efficiency. In a spread-spectrum system, bandwidth efficiency is usually not of prime concern. Rather, the use of QPSK is motivated by the fact that it is less sensitive to some types of interference (jamming).

In an FH/MFSK system, the PN sequence makes the carrier hop over a number of frequencies in a pseudo-random manner, with the result that the spectrum of the transmitted signal is spread in a sequential manner.

Naturally, the direct-sequence and frequency-hop spectrum-spreading techniques may be employed in a single system. The resulting system is referred to as *hybrid* DS/FH *spread-spectrum system*. The reason for seeking a hybrid approach is that advantages of both the direct-sequence and frequency-hop spectrum-spreading techniques are realized in the same system.

A discussion of spread-spectrum communications would be incomplete without some reference to jammer waveforms. The jammers encountered in practice include the following types:

1. *The barrage noise jammer*, which consists of band-limited white Gaussian noise of high average power. The barrage noise jammer is a brute-force jammer that does not exploit any knowledge of the antijam communication system except for its spread bandwidth.
2. *The partial-band noise jammer*, which consists of noise whose total power is evenly spread over some frequency band that is a subset of the total spread bandwidth. Owing to the smaller bandwidth, the partial-band noise jammer is easier to generate than the barrage noise jammer.
3. *The pulsed noise jammer*, which involves transmitting wideband noise of power

$$J_{\text{peak}} = \frac{J}{p}$$

for a fraction p of the time, and nothing for the remaining fraction $1 - p$ of the time. The average noise power equals J.
4. *The single-tone jammer*, which consists of a sinusoidal wave whose frequency lies inside the spread bandwidth; as such, it is the easiest of all jamming signals to generate.
5. *The multitone jammer*, which is the tone equivalent of the partial-band noise jammer.

In addition to these five, many other kinds of jamming waveforms occur in practice. In any event, there is no single jamming waveform that is worst for all spread-spectrum

systems, and there is no single spread-spectrum system that is best against all possible jamming waveforms.

NOTES AND REFERENCES

1. The definition of spread-spectrum modulation presented in the Introduction is adapted from Pickholtz, Schilling, and Milstein (1982). This paper presents a tutorial review of the theory of spread-spectrum communications.

   For introductory papers on the subject, see Viterbi (1979), and Cook and Marsh (1983). For books on the subject, see Dixon (1984), Holmes (1982), Ziemer and Peterson (1985, pp. 327–649), Cooper and McGillem (1986, pp. 269–411), and Simon, Omura, Scholtz, and Levitt (1985, Volumes I, II, and III). The three-volume book by Simon et al. is the most exhaustive treatment of spread-spectrum communications available in the open literature. The development of spread-spectrum communications dates back to about the mid-1950s. For a historical account of these techniques, see Scholtz (1982). This latter paper traces the origins of spread-spectrum communications back to the 1920s. Much of the historical material presented in this paper is reproduced in Chapter 2, Volume I, of the book by Simon et al.

   The book edited by Tantaratana and Ahmed (1998) includes introductory and advanced papers on wireless applications of spread-spectrum modulation. The papers are grouped into the following categories: spread-spectrum technology, cellular mobile systems, satellite communications, wireless local area networks, and global positioning systems (GPS).

2. For further details on maximal-length sequences, see Golomb (1964, pp. 1–32), Simon, Omura, Scholtz, and Levitt (1985, pp. 283–295), and Peterson and Weldon (1972). The last reference includes an extensive list of polynomials for generating maximal-length sequences; see also Dixon (1984). For a tutorial paper on pseudo-noise sequences, see Sarwate and Pursley (1980).

3. Table 7.1 is extracted from the book by Dixon (1984, pp. 81–83), where feedback connections of maximal-length sequences are tabulated for shift-register length m extending up to 89.

4. For detailed discussion of the synchronization problem in spread-spectrum communications, see Ziemer and Peterson (1985, Chapters 9 and 10) and Simon et al. (1985, Volume III).

5. The original papers on Gold sequences are Gold (1967, 1968). A detailed discussion of Gold sequences is presented in Holmes (1982).

PROBLEMS

Pseudo-Noise Sequences

7.1 A pseudo-noise (PN) sequence is generated using a feedback shift register of length $m = 4$. The chip rate is 10^7 chips per second. Find the following parameters:

(a) PN sequence length.

(b) Chip duration of the PN sequence.

(c) PN sequence period.

WEEK 15

7.2 Figure P7.2 shows a four-stage feedback shift register. The initial state of the register is 1000. Find the output sequence of the shift register.

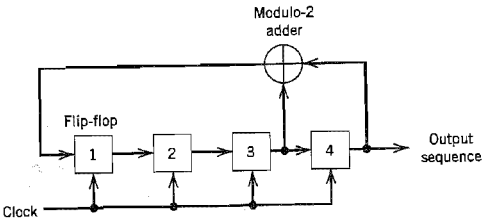


FIGURE P7.2

7.3 For the feedback shift register given in Problem 7.2, demonstrate the balance property and run property of a PN sequence. Also, calculate and plot the autocorrelation function of the PN sequence produced by this shift register.

7.4 Referring to Table 7.1, develop the maximal-length codes for the three feedback configurations [6, 1], [6, 5, 2, 1], and [6, 5, 3, 2], whose period is $N = 63$.

7.5 Figure P7.5 shows the modular multitap version of the linear feedback shift-register shown in Figure 7.4*b*. Demonstrate that the PN sequence generated by this scheme is exactly the same as that described in Table 7.2*b*.

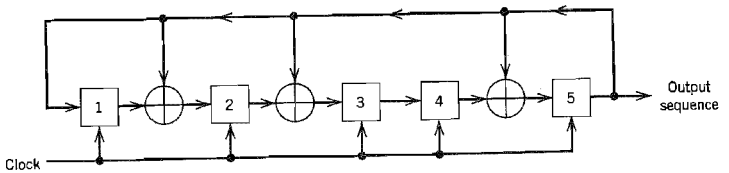


FIGURE P7.5

Direct Sequence/Phase-Shift Keying System

7.6 Show that the truth table given in Table 7.3 can be constructed by combining the following two steps:

(a) The message signal $b(t)$ and PN signal $c(t)$ are added modulo-2.

(b) Symbols 0 and 1 at the modulo-2 adder output are represented by phase shifts of 0 and 180 degrees, respectively.

7.7 A single-tone jammer

$$j(t) = \sqrt{2J}\cos(2\pi f_c t + \theta)$$

is applied to a DS/BPSK system. The N-dimensional transmitted signal $x(t)$ is described by Equation (7.16). Find the $2N$ coordinates of the jammer $j(t)$.

7.8 The processing gain of a spread-spectrum system may be expressed as the ratio of the spread bandwidth of the transmitted signal to the despread bandwidth of the received signal. Justify this statement for the DS/BPSK system.

7.9 A direct-sequence spread binary phase-shift keying system uses a feedback shift register of length 19 for the generation of the PN sequence. Calculate the processing gain of the system.

7.10 In a DS/BPSK system, the feedback shift register used to generate the PN sequence has length $m = 19$. The system is required to have an average probability of symbol error due to externally generated interfering signals that does not exceed 10^{-5}. Calculate the following system parameters in decibels:

(a) Processing gain.

(b) Antijam margin.

7.11 In Section 7.5, we presented an analysis on the signal-space dimensionality and processing gain of a direct sequence spread-spectrum system using binary phase-shift keying. Extend the analysis presented therein to the case of such a system using quadriphase-shift keying.

Frequency-Hop Spread Spectrum

7.12 A slow FH/MFSK system has the following parameters:

Number of bits per MFSK symbol = 4

Number of MFSK symbols per hop = 5

Calculate the processing gain of the system.

7.13 A fast FH/MFSK system has the following parameters:

Number of bits per MFSK symbol = 4

Number of hops per MFSK symbol = 4

Calculate the processing gain of the system.

Computer Experiments

7.14 Consider two PN sequences of period $N = 63$. One sequence has the feedback taps [6, 1] and the other sequence has the feedback taps [6, 5, 2, 1], which are picked in accordance with Table 7.1.

(a) Compute the autocorrelation function of these two sequences, and their cross-correlation function.

(b) Compare the cross-correlation function computed in part (a) with the cross-correlation function between the sequence [6, 5 2, 1] and its mirror image [6, 5, 4, 1].

Comment on your results.

7.15 (a) Compute the partial cross-correlation function of a PN sequence with feedback taps [5, 2] and its image sequence defined by the feedback taps [5, 3].

(b) Repeat the computation for the PN sequence with feedback taps [5, 2] and the PN sequence with feedback taps [5, 4, 2, 1].

(c) Repeat the computation for the PN sequence with feedback taps [5, 4, 3, 2] and the PN sequence with feedback taps [5, 4, 2, 1].

The feedback taps [5, 2], [5, 4, 3, 2], and [5, 4, 2, 1] are possible taps for a maximal-length sequence of period 31, in accordance with Table 7.1.