CHAPTER 10

LOW EARTH ORBIT AND NON-GEOSTATIONARY SATELLITE SYSTEMS

For 6 years after the launch of *Sputnik 1* on the night of October 4/5 1957, all of the space vehicles launched were either inserted into low earth orbit (LEO) or they were small scientific research vehicles dispatched on missions to the earth's moon or to the inner planets: Mars and Venus. The primary reason for the use of LEO was the generally small *throw mass* of the launchers then in use. The throw mass of a launcher is the mass of the spacecraft that is ejected from the spent rocket. The throw mass includes both the payload for the mission and the spacecraft bus system within which it will function. It also includes any additional rocket motors and fuel that will be needed to boost the spacecraft into the final trajectory and/or to maintain that trajectory. While the throw mass of the launchers used by the former USSR in that period was significantly higher than that of the U.S. launchers, political reasons dictated their capabilities be used in noncommercial areas. It was therefore left to the United States to open up the geostationary (Clarke) orbit[1] in 1963 with Syncom 1 and with it, the commercialization of space. The Clarke orbit, more usually called the geostationary earth orbit (GEO) or geostationary satellite orbit (GSO), is a unique resource that has enabled the generation of many billions of dollars in revenues per year from communications satellites and the associated launchers. Communications satellites and their launchers were the only commercial space ventures in the twentieth century that had any significant return on investment. This may change with the new generation of non-geostationary satellite orbit satellite constellations currently under development, in deployment, and in operation for a variety of commercial ventures, although the first ventures have unfortunately been conspicuous failures.

The terms low earth orbit (LEO) and medium earth orbit (MEO) are generally used for specific orbit altitude ranges, for reasons that we will see later. LEO satellites are confined between an upper orbit altitude of about 1500 km and a lower orbit altitude dictated by atmospheric drag (generally around 500 km). MEO satellites have a lower orbit altitude of around 1500 km and an upper bound set by the GEO altitude of around 36,000 km. Most MEO systems, however, orbit in the 10,000 to 15,000 km range.

LEO and MEO satellites—now generally referred to as non-geo-stationary orbit (NGSO) satellites—have been used in a variety of roles. From an era in the late 1950s when every launch made front-page news we have now become somewhat blasé about satellites: they have become part of everyday life, much like computers and the Internet. NGSO satellites brought us the first communications satellite (SCORE), the first pictures of our cloud cover for weather forecasting (TIROS), the first navigation aids in space (TRANSIT), the first live television pictures across oceans (TELSTAR), the first Geographic Information Systems pictures of the earth (SPOT), the first infrared, ultraviolet, and X-ray view of the universe from outside the earth's atmosphere and, of course, the first manned missions (Vostok and Mercury). Each of these missions has been succeeded by more complex satellites with more advanced capabilities: soon the International Space Station (ISS) will assume scientific missions accompanied by one, or more, free-flying modules and, for the first time, watchers on the earth with their naked eye in daytime may be able to see a satellite as it passes overhead. Already, in early 2002, the ISS had an orbital mass in excess of 250,000 lb and dimensions close to a Boeing 747. As the satellite missions became more complex, the requirements for the specific orbits became more precise. Some satellites have to be very close to the earth, some in highly elliptical orbits, and yet others in orbits with a plane that matches the view angle to the sun. This chapter reviews the different earth orbits available and what missions may use them to advantage.

10.1 INTRODUCTION

The geostationary orbit has been the preferred orbit for satellite communication systems for 35 years, and is likely to continue to be the orbit that provides most of the revenue for satellite system operators. The reason is simple: more bits can be sent per dollar of capital investment when a satellite is in a geostationary orbit than in any other orbit. This was realized quite early in the development of satellite communications, and Intelsat, which was the first provider of commercial satellite systems, developed a series of geostationary satellites, beginning in 1965 with *Early Bird* (INTELSAT I). Commercial and national satellite systems followed in the 1970s and 1980s, all using GEO satellites. Direct-to-home (DTH) satellite television broadcasting, one of the most successful applications for satellite communication systems, also requires GEO satellites so that customers can use small fixed dish antennas. In such a DBS-TV system, the major investment is in earth stations, not in the satellite. Ten million earth stations bought for $250 each, for example, cost $2.5 billion, well in excess of the cost of a cluster of GEO DBS-TV satellites.

There are some specialized applications that require non-geostationary satellites. Surveillance of the earth's surface, for both military data gathering and earth resources applications, requires satellites in low earth orbit that cover the entire surface of the earth. Satellites providing global navigation, such as the Global Positioning System (GPS) constellation, must utilize orbits that place the satellites in widely spaced positions in the sky, as seen by the receiver. Some of the satellites can be in GEO, but most must be in inclined orbits with an even distribution over the earth's surface. GPS uses 24 satellites in orbits with an altitude of 20,000 km and an inclination of 55°.

Mobile satellite communication systems demand an earth station with a low gain antenna that has a near omnidirectional pattern. A GEO satellite used for communication with a satellite telephone that is handheld, like a cellular telephone, requires a very large antenna with hundreds of beams to achieve a very high gain. The high gain satellite antenna is needed to compensate for the low gain of the antenna employed by the user's telephone handset. An alternative to a GEO satellite with a high gain antenna is a LEO or MEO satellite constellation with a smaller multibeam antenna. Because the satellite is not geostationary, a large number of satellites is required to maintain continuous coverage. The Iridium system used 66 satellites in LEO, for example, to provide continuous global coverage.

Building, launching, and maintaining a constellation of communication satellites in low earth orbit is expensive. When low earth orbit satellite constellations were first proposed for mobile satellite services, the satellites were envisaged to be small, simple, and low cost compared with GEO satellites. Early estimates for the cost of the Iridium system, for example, were between $1 billion and $2 billion. As the development of the LEO systems progressed, the satellites became more and more complex and their cost steadily increased, becoming comparable to the cost of GEO satellites. The satellites used by the ICO global system (now called New ICO), for example, are actually modified versions of a large GEO satellite, the Hughes (Boeing) 601 design. Since any LEO or MEO system requires many more satellites than a GEO system serving the same region, the cost of the LEO or MEO system will exceed the cost of the equivalent GEO system.

The Iridium system, with 66 satellites in low earth orbit, eventually cost over $5 billion, compared with a typical cost of $250 million to launch and maintain a large GEO satellite. Iridium failed as a commercial venture because the final cost greatly exceeded initial projections, and the system was unable to attract a sufficient number of customers quickly enough. Debt repayments on the high capital cost of the system came due before the customer base had built to a large enough size to provide substantial revenue. However, analysis of the cost per bit transmitted through an Iridium satellite shows that it is much higher than the cost per bit for a GEO satellite, and any LEO system must therefore be able to offer considerable advantages to its customers over that of an equivalent GEO system if it is to succeed commercially. It remains to be seen whether the other non-geostationary mobile satellite systems can succeed where Iridium failed. As the twenty-first century starts, prospects do not look bright for NGSO systems with Iridium, Orbcomm, Globalstar, and ICO filing for bankruptcy protection. ICO has emerged from the Chapter 11 filing as New ICO, but is still struggling with identifying its mission. Initially conceived as a mobile satellite system provider, emphasis has moved to the provisioning of Internet-like service to mobile customers as a preliminary to Teledesic.

This chapter discusses a number of applications and satellite systems that are not in GEO orbit, beginning with those in simple, circular, equatorial orbits; moving through simple inclined orbits to those with high eccentricity; and then reviewing those that take advantage of specific attributes of their orbit for observations (sun synchronous orbits) or the provisioning of navigation services through half-sidereal periodic orbits (GPS). The so-called inclined orbit GEO satellites are not discussed in this chapter. These satellites, once fully stationary GEO satellites, have had their in-orbit operational life extended by removing station keeping in the N–S direction while maintaining E–W station keeping so that the average subsatellite point remains nominally the same. Such inclined-orbit operation was first started after an unusual run of launch vehicle failures in the 1986 time frame when every single type of commercial satellite launcher failed (including the tragic loss of the space shuttle Challenger). The up to 2-year hiatus in some satellite replenishment programs forced inclined orbit operation of GEO satellites on all service providers. Currently, such

SIDEBAR

The first spacecraft launches relied on terrestrial radar tracking and guidance commands transmitted from the ground. This, and the relatively crude control capabilities of the rockets themselves, dictated relatively wide error bounds for the intended orbit. Indeed, achieving orbit in those early days—any orbit—was declared a success! Rapid advances in rocketry, which included the ability for multiple restarts of high-energy upper stage engines, and the inclusion of sophisticated on-board guidance computers, quickly enabled spacecraft mission planners to design with some confidence orbits that were mission-specific. That is, the mission could not be successful *unless* the designed orbit was achieved within the specified tolerance. In some cases, the mission was for a single spacecraft (such as a meteorological satellite) while, in others, a constellation of spacecraft would be required to achieve the mission goals. In all cases, careful analysis of the mission goals led to the selection of a particular orbit altitude, ellipticity, and inclination and system architecture (number of satellites, number of planes, spacing of satellites within the plane, connectivity, etc.). Quite often, tight launch windows were also dictated—specific time periods when the launches had to be executed.

problems are avoided by designing the orbital maneuvering life (OML) to be many years longer then the orbital design life (ODL). This aspect is discussed in Chapter 2.

In the sections that follow, we will examine the parameters that need to be determined in the selection of an orbit that will achieve given mission goals. Only earth orbit missions are considered; spacecraft missions requiring escape velocity from the earth are beyond the scope of this book.

10.2 ORBIT CONSIDERATIONS

Once in orbit, the motion of a satellite is determined by orbital mechanics, as discussed in Chapter 2, *Orbital Mechanics and Launchers*. However, while the satellite moves in such a way as to balance centrifugal and centripetal forces, the earth is also in motion beneath it. As well as rotating once a sidereal day, the earth also moves around the sun; and the solar system, with the sun at its center, is orbiting around the center of the home galaxy, the Milky Way. There is therefore a complex relationship between the various motions of the natural and artificial bodies. How many of these need to be considered simultaneously will depend on the design goals of the satellite system. A satellite designed to observe the earth's surface will not need to know where the stars are at any particular time, but the location of the local star, the sun, may be important if the satellite needs to use sunlight to illuminate its coverage region on the surface of the earth. On the other hand, a satellite designed to observe background thermal radiation levels of deep space in the infrared band will need to know the position of each of the neighboring planets. Should the telescope of the satellite inadvertently point toward one of these planets, the temperature viewed would not reflect that of the true background radiation level. In the sections that follow, we will review all of the different NGSO orbits that have been used for scientific, military, and commercial satellite missions. The simplest NGSO orbit is an equatorial orbit.

Equatorial Orbits

Equatorial orbits lie exactly in the plane of the geographical equator of the earth. That is, the orbital path lies directly above the equator at all times. In order to take advantage of the 0.45 km/s eastward rotational velocity of the earth, most satellites are launched toward the east into a *prograde* orbit. A westerly directed orbit is called a *retrograde* orbit. A satellite in an eastwardly directed equatorial orbit will have two periods: a real orbital period that is referenced to inertial space (the galactic background) and an apparent orbital period that is referenced to a stationary observer on the surface of the earth. The real orbital period, denoted here as T hours, is given by Eq. 2.6. The apparent orbital period to the observer on the equator will be P hours where

$$P = (24T)/(24 - T) \text{ hours} \tag{10.1}$$

To be exact, 23.9344 h, one sidereal day, should be used in place of 24 h in Eq. (10.1). Table 10.1 (from reference 2) illustrates the difference between P and T for a number of orbital altitudes and elevation angles. It also shows the time the satellite is visible to the observer, neglecting atmospheric refraction and assuming the satellite can be tracked down to 0°, that is, right down to the horizon. Other implications of the observing time are considered in more detail in Section 10.3, *Coverage and Frequency Considerations*.

The plane of a satellite's orbit must be in the plane of the equator for the satellite to be in equatorial orbit. This can be achieved by launching the satellite in one of two ways. The first launch method is to locate the launch site on the equator and to launch

TABLE 10.1 Orbital Periods and Observing Time

Orbital height (km)	Orbital period		Observing Time (hours)
	True (hours)	Apparent (hours)	
500	1.408	1.496	0.183
1,000	1.577	1.688	0.283
5,000	1.752	1.890	0.587
10,000	5.794	7.645	2.894
35,786	23.934	∞	∞

Source: From Table 1.1 in reference 2.

the spacecraft toward the east along the equatorial plane. The second method is to launch the satellite into an inclined orbit and to execute a maneuver either during the launch trajectory or when the satellite is in an inclined orbit that changes the plane of the initial orbit so that the final orbit is in the plane of the equator. Removing the inclination from the orbit, so that the satellite orbits exactly over the equator, requires significant energy, particularly if the launch site is well removed from the equator. The first two sites from which orbital flights were made, Cape Canaveral in the United States and Baikonur in Kazakhstan, were not close to the equator (approximately 28° N and 46° N, respectively). In addition, the early launch vehicles lacked the ability to alter the trajectory significantly during launch. The first artificial earth satellites were therefore placed into inclined orbits, that is, the planes of the orbits were inclined to the equatorial plane.

Inclined Orbits

There are advantages and disadvantages to inclined orbits, depending on the mission goals and the data recovery requirements. The greater the inclination of the orbit is, the larger the surface area of the earth that the satellite will pass over at some time in its flight. Figure 10.1 illustrates this for a LEO satellite.

In Figure 10.1*b*, the inclined orbit will take the spacecraft, at one time or another, over the earth's entire surface that lies approximately between the latitudes given by ± the orbital inclination. For example, an orbit with an inclination of 30° will cover all regions that lie approximately between latitudes 30° north and 30° south. The superior coverage of the earth with an inclined orbit satellite is counterbalanced by the disadvantage that the master control station (MCS) will not be able to communicate directly with the satellite on every orbit as with an equatorial orbit satellite. A LEO satellite orbits the earth with a period of 90 to 100 min and, for an inclined orbit satellite, the earth will have rotated the master control station out of the path of the satellite on the next pass over the same side of the earth. Depending on the quantity of data that need to be passed to the MCS, or if real-time communications are required continuously, a system architecture that employs multiple satellites will need to be considered.

The simplest, and lowest cost, solution to pass data between an inclined orbit satellite and an MCS is to design the satellite to store the data acquired over many orbits (when it is out of sight of the MCS) and then, when it passes within radio range of the MCS, to dump the data rapidly to the MCS. This is called *store-and-forward* and it is one of the capabilities of some LEO systems, including Orbcomm satellites[3]. It was also the technique used for the very first communications satellite, Project SCORE, in December 1958. In the

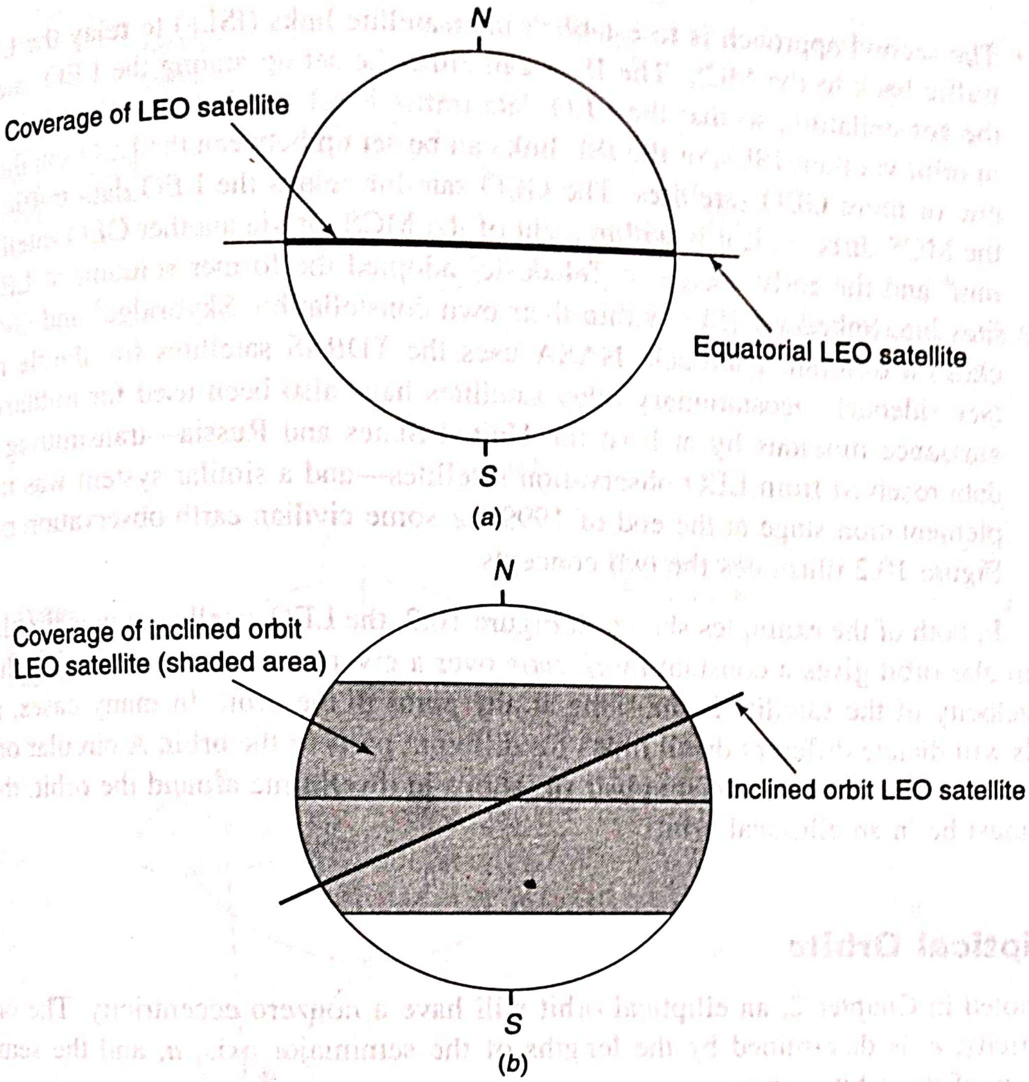


FIGURE 10.1 (*a*) Coverage of an equatorial orbit LEO satellite. The LEO satellite is in an equatorial LEO orbit and so it will only pass over the equator. The coverage of the equatorial LEO satellite will therefore be limited to a swathe of the earth close to the equator, determined by the height of the orbit and the beamwidth of the satellite's antenna. In this example, the orbit is assumed to be circular and the antenna beamwidth has been ignored. (*b*) Coverage of an inclined orbit LEO satellite. The LEO satellite is in an orbit that is inclined at approximately 40° to the equator. The satellite will therefore pass over, at one time or another, all regions of the earth between latitudes 40° N and 40° S of the equator. The coverage of the inclined orbit LEO satellite will therefore be a swathe of the earth between about ±40° of the equator, determined by the height of the orbit and the beamwidth of the satellite's antenna. In this example, the orbit is assumed to be circular and the antenna beamwidth has been ignored. Note: The higher the orbit and the greater the inclination, the further the satellite's total coverage will reach.

Orbcomm system, if a user on the ground is unable to establish contact via an Orbcomm satellite to a gateway earth station (GES) in the Orbcomm system, a "GlobalGram®" may be left stored within the satellite for later transmission to the GES when it comes into view of the satellite. The downlink transmission rate must be high enough to enable all of the stored messages in a LEO satellite to be sent to the MCS in the period when it is within range of the satellite. If a continuous, real-time connection is required between a LEO satellite and the MCS, there are only two approaches that can be used.

- The first approach is to locate control stations around the world so that the LEO satellite is never out of sight of at least one of the control stations. Terrestrial or GEO satellite connections are then established between the many control stations and the MCS to bring the LEO data back to the MCS in real time.

- The second approach is to establish intersatellite links (ISLs) to relay the LEO data traffic back to the MCS. The ISLs can either be set up among the LEO satellites in the constellation, so that the LEO data traffic is relayed between the LEO satellites in orbit via their ISLs, or the ISL link can be set up between the LEO satellite(s) and one or more GEO satellites. The GEO satellite relays the LEO data traffic back to the MCS directly, if it is within sight of the MCS, or via another GEO satellite. Iridium[4] and the early design of Teledesic[5] adopted the former solution of LEO satellites interlinked via ISLs within their own constellation. Skybridge[6] and Globalstar[7] chose a different approach. NASA uses the TDRSS satellites for shuttle missions (see sidebar). Geostationary relay satellites have also been used for military reconnaissance missions by at least the United States and Russia—transmitting onward data received from LEO observation satellites—and a similar system was in its implementation stage at the end of 1999 for some civilian earth observation missions. Figure 10.2 illustrates the two concepts.

In both of the examples shown in Figure 10.2, the LEO satellite is in a circular orbit. A circular orbit gives a constant *dwell time* over a given coverage region since the angular velocity of the satellite is the same at any point in the orbit. In many cases, mission goals will dictate different dwell times for different parts of the orbit. A circular orbit will not achieve this result. To accomplish variations in dwell time around the orbit, the satellite must be in an elliptical orbit.

Elliptical Orbits

As noted in Chapter 2, an elliptical orbit will have a nonzero eccentricity. The orbit eccentricity, e, is determined by the lengths of the semimajor axis, a, and the semiminor axis, b, of the orbit ellipse

$$e^2 = 1 - (b^2/a^2) \tag{10.2}$$

Alternatively, if R_a is the distance between the center of the earth and the apogee point of the orbit and R_p is the distance between the center of the earth and the perigee point, the eccentricity is

$$e = (R_a - R_p)/(R_a + R_p) \tag{10.3}$$

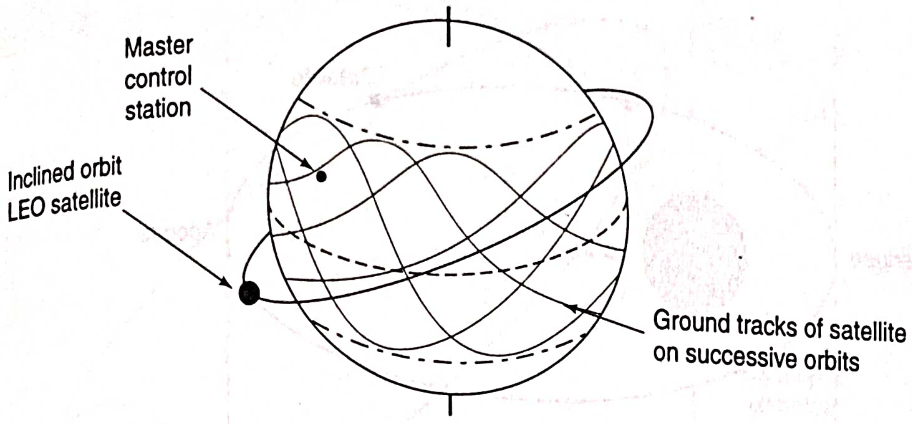
Figure 10.3 illustrates the geometry of Eq. (10.3).

In Eqs. (10.2) and (10.3), if the orbit is exactly circular, $a = b$ and $R_a = R_p$, and the eccentricity reduces to zero. In general, no orbit is truly circular for a variety of reasons,

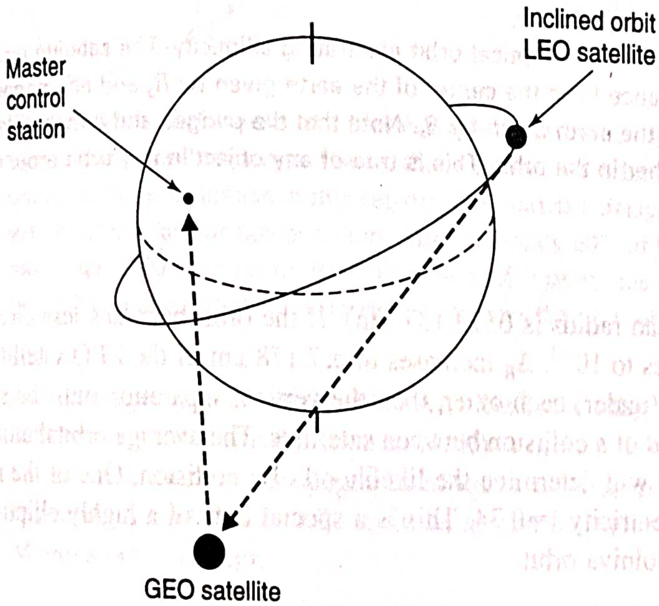
SIDEBAR

NASA built and operated a number of relay stations for the Mercury, Gemini, and Apollo programs. In none of these missions was the manned spacecraft ever out of real-time contact with the Manned Spaceflight Center in Houston, United States (which acted as the MCS in this case) except when the Apollo craft was behind the moon or in the re-entry phase where ionized plasma caused a radio blackout for all spacecraft.

Communication with the Space Shuttle is maintained using the Tracking and Data Relay Satellite System (TDRSS). Several TDRSS satellites in geostationary orbit relay data from the Shuttle to several earth stations around the world that then send the data to NASA's MCS for manned space flight in Houston, Texas.



(a)



(b)

FIGURE 10.2 (*a*) Store-and-forward concept. In this LEO application, the satellite stores information it has gathered while orbiting the earth and, once within range of the master control station, it downloads the stored data. The (uplinked) data storage rate is usually low, a few kbit/s at most, while the download is at a much higher rate due to the small time the satellite has available when it is within range of the master control station. (*b*) Real-time data transfer via a GEO satellite. In this approach, the LEO satellite can transfer data in real time via the GEO satellite to the master control station whenever it can "see" the GEO satellite. If there were a number of GEO satellites equipped with intersatellite links (ISLs) distributed around the geostationary orbit, then the LEO satellite need never be out of real-time contact with the master control station.

but eccentricity values of 10^{-3} or less can be considered to correspond to circular orbits for all practical purposes. The eccentricity is another way of describing the variation in the radius of the orbit. If R_{av} is the average radius of an orbit from the center of the earth, then the variation, Δ_R, in the orbital radius, is given by[8]

$$\Delta_R = \pm e R_{av} \tag{10.4}$$

For a geostationary satellite (R_{av} = 42,164.17 km) with an eccentricity of 10^{-4}, Δ_R will be ±4.2 km. For a LEO constellation with a circular orbit of approximately 800 km above the earth, with each LEO satellite having an eccentricity of 10^{-4}, Δ_R will be ±0.7178 km

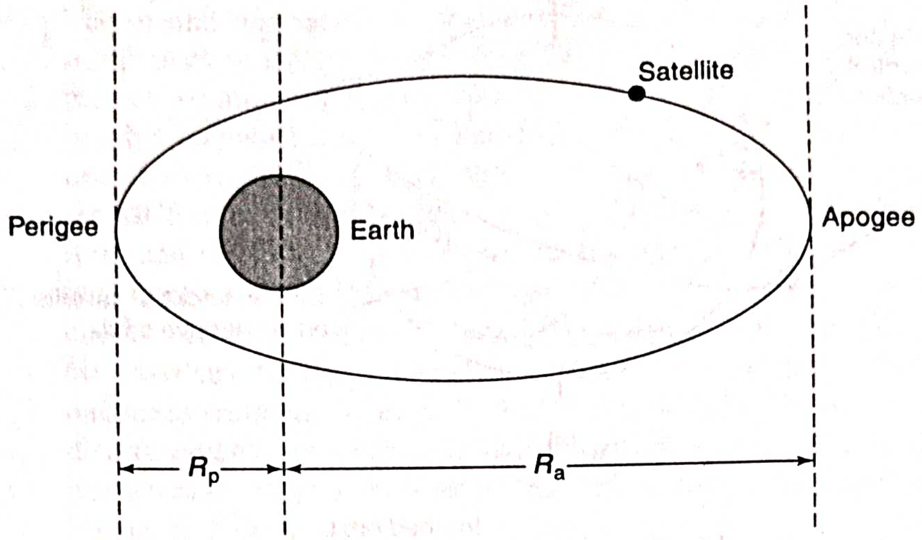


FIGURE 10.3 Schematic of an elliptical orbit illustrating ellipticity. The satellite orbits the earth with a perigee distance from the center of the earth given by R_p and an apogee distance from the center of the earth given by R_a. Note that the perigee and apogee are *always* exactly opposite each other in the orbit. This is true of any object in any orbit around any other body.

(assuming the earth mean radius is 6378.137 km). If the orbit becomes less circular and the eccentricity increases to 10^{-3}, Δ_R increases to ± 7.178 km. If the LEO satellites in the constellation pass over (under) each other, then the vertical separation must be sufficient to prevent any likelihood of a collision between satellites. The average orbital altitude and eccentricity of the orbit will determine the likelihood of a collision. One of the more famous orbits has an eccentricity ≈ 0.74. This is a special case of a highly elliptical orbit (HEO) known as the Molniya orbit.

Molniya Orbit

The former Soviet Union had a difficult communications design problem. Much of the landmass is in far northern latitudes. Archangel, the port on the White Sea, is close to latitude 60° N; immense tracts of Siberia lie inside the Arctic Circle. To compound the problem further, the country was spread across 11 time zones: it was the largest country in the world (and Russia still is). The signals from a geostationary satellite can reach well inside the Arctic Circle if operations at elevation angles below 5° are permitted, but a single GEO satellite cannot reach that far north over 11 time zones simultaneously. A new type of orbit was required to provide good communications coverage over the former USSR. What transpired was the Molniya system.

The first Molniya satellite was launched in April 1965 and it gave its name to both the system of satellites and to the unique orbit. The word Molniya means *flash of lightning* in Russian. The apogee of the Molniya orbit is at an altitude of 39,152 km and the perigee is at an altitude of 500 km. The orbital period is 11 h and 38 min and the orbital inclination is 62.9°. This combination of apogee, perigee, and inclination ensures that the ground track of the Molniya orbit repeats every other orbit. That is, if the orbit passes exactly over Moscow on orbit one, it will do so again on orbit three, five, seven, nine, and so on. Figure 10.4 illustrates the orbit geometry.

Two Molniya orbits, with the planes of the orbits separated by 180°, will thus provide coverage over the extreme latitudes of Russia for 24 hours per day using two satellites, correctly phased—one in each of the two Molniya orbits. When one of the satellites

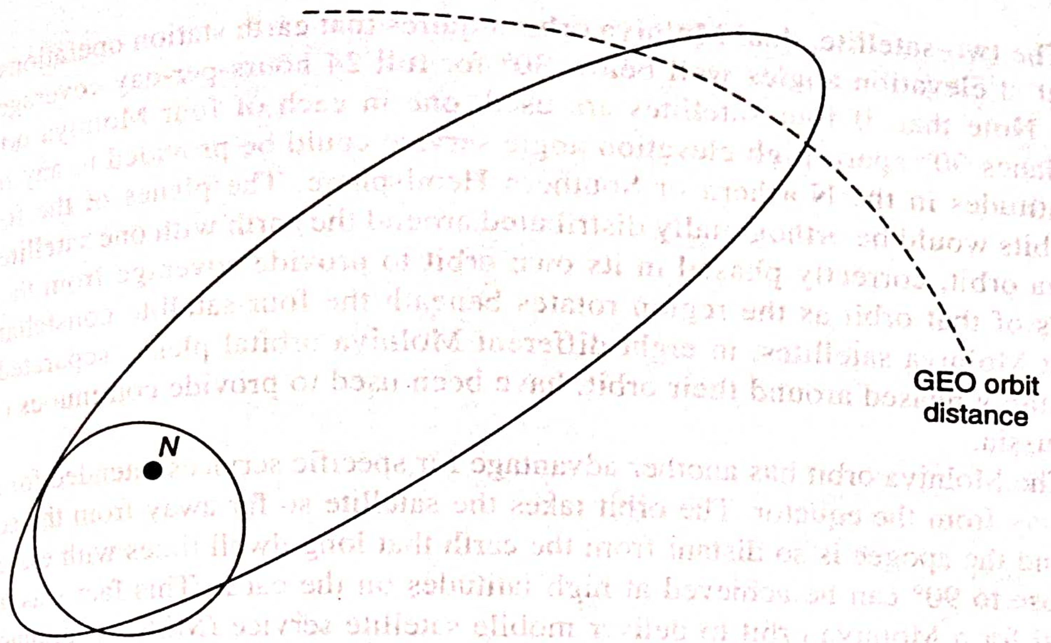


FIGURE 10.4 Schematic of a Molniya orbit. In this example, the trajectory is configured to have a large dwell time over the northern part of the orbit so that it can serve a country that has most of its landmass in this region. This was the design adopted for the original Molniya system of the former Soviet Union. Approximately 60% of this Molniya orbit, which stretches more than 3000 km beyond the height of a GEO orbit, has good look angles for latitudes between 30° N and 90° N. This translates to more than 6 h of the 11 h 38 min orbital period.

is at its apogee over Russia in Molniya orbit 1, the other satellite will also be at its apogee somewhere over North America in Molniya orbit 2. By the time the second satellite has moved once more to its apogee in Molniya orbit 2, the earth will have revolved half a turn under it and Russia will again be spread beneath it. Figure 10.5 illustrates the dual Molniya orbit concept.

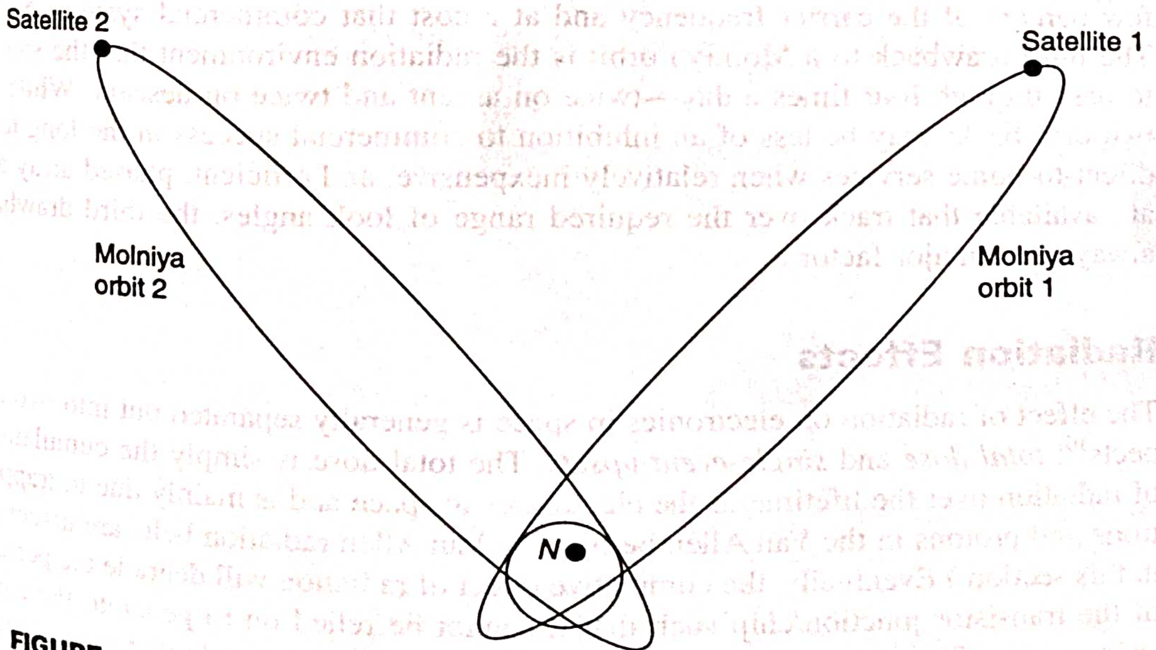


FIGURE 10.5 Schematic of an operational Molniya system. Satellite 1 in Molniya orbit 1 is providing service over Russia at close to its apogee while the second satellite is also close to its apogee in Molniya orbit 2. Molniya orbits I and 2 are separated by 180° in their orbital planes. By the time satellite 2 has moved around its orbit once and back to its apogee (a period of about 12 hours), the earth will have rotated about 180° and the second satellite will be over Russia.

The two-satellite, dual Molniya orbit requires that earth station operations be carried out at elevation angles well below 30° for full 24 hours-per-day coverage of one region. Note that, if four satellites are used, one in each of four Molniya orbits that have planes 90° apart, high elevation angle service could be provided to any region at high latitudes in the Northern or Southern Hemisphere. The planes of the four Molniya orbits would be orthogonally distributed around the earth with one satellite in each Molniya orbit, correctly phased in its own orbit to provide coverage from the apogee sections of that orbit as the region rotates beneath the four-satellite constellation. Up to eight Molniya satellites, in eight different Molniya orbital planes separated by 45° and suitably phased around their orbit, have been used to provide continuous coverage over Russia.

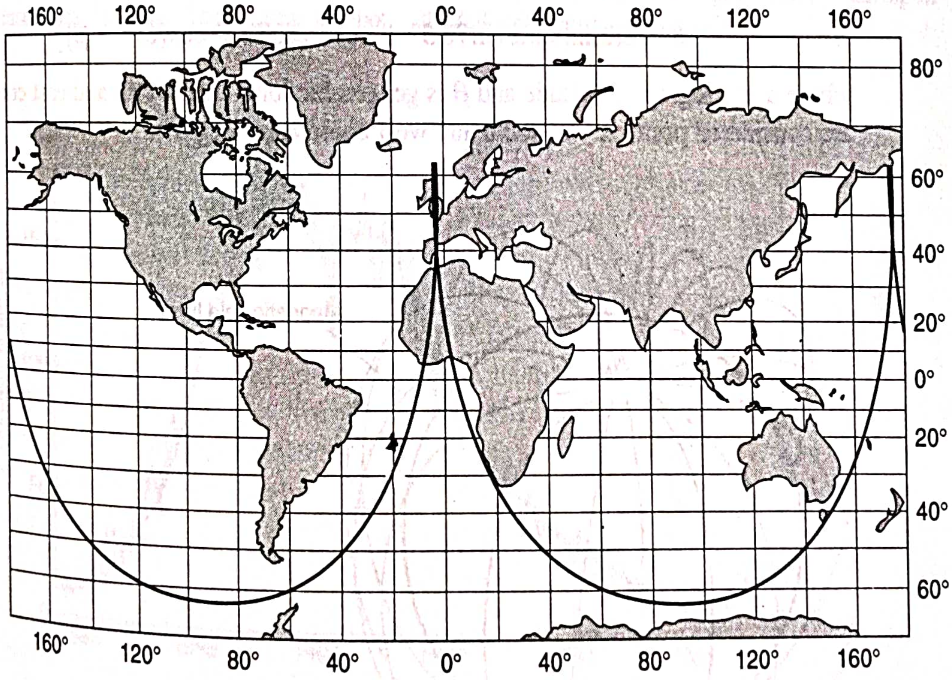
The Molniya orbit has another advantage for specific services intended for latitudes well away from the equator. The orbit takes the satellite so far away from the equatorial plane and the apogee is so distant from the earth that long dwell times with elevation angles close to 90° can be achieved at high latitudes on the earth. This fact was used in a proposal for a Molniya orbit to deliver mobile satellite service (MSS) to automobiles. A view of the earth, with the plot of the orbit track, abstracted from this proposal[9] is shown in Figure 10.6.

The Molniya orbit, in addition to the long delay time associated with the communications range when at apogee and the lack of continuous 24-h contact with a single spacecraft from a fixed coverage, also suffers from three drawbacks that increase the overall end-to-end costs. The first is the requirement to track the spacecraft. The second is the need to switch communications to the other Molniya satellite—rather like a mobile radio handoff situation—when the first goes out of coverage as the other comes into coverage. Due to the wideband nature of the traffic and the large angular separation between successive Molniya satellites as seen from one earth station, this requires two reflector antennas at each site. At the end of the twentieth century, phased array antennas still could not provide accurate coincident tracking of both transmit and receive beams simultaneously well away from the (unsteered) electrical boresight over bandwidths that exceed a few percent of the carrier frequency and at a cost that commercial systems can accept. The third drawback to a Molniya orbit is the radiation environment that the satellite has to pass through four times a day—twice on ascent and twice on descent. While the first two drawbacks may be less of an inhibition to commercial success in the long term with direct-to-home services when relatively inexpensive, and efficient, phased array antennas are available that track over the required range of look angles, the third drawback will always be a major factor.

Radiation Effects

The effect of radiation on electronics in space is generally separated out into two main aspects[10]: *total dose* and *single-event upsets*. The total dose is simply the cumulative effect of radiation over the lifetime of the electronics in space and is mainly due to trapped electrons and protons in the Van Allen belts. (The Van Allen radiation belts are discussed later in this section.) Eventually, the cumulative effect of radiation will degrade the performance of the transistor junction/chip such that it cannot be relied on to generate the correct responses, etc. This is particularly harmful in the computer elements that control the operation of the satellite and the payload. Single-event upsets are caused by heavy ions ejected from the sun, usually protons, impacting the circuitry at a critical point such that they deposit enough charge to induce an energy (bit) flip, that is, change an open circuit to a closed circuit, create a logical one instead of a logical zero, etc. These single-event upsets are

(a)



(b)

FIGURE 10.6 View from above the Molniya orbit apogee showing the ground track[8] (*a*) View from the apogee point of a Molniya orbit positioned at almost 0° longitude when at apogee. (*b*) Ground track of the Molniya orbit shown in Figure 10.6*a*. Note the two apogees in the orbit, one over close to 0° longitude and the other at close to 180°. The apogee occurs at a high latitude, from which the elevation angles are well above 70° over quite a large region. With these high elevation angles, blockage of buildings would be minimized and thus allow relatively high availability for an MSS system operating to automobiles in most cities. This proposal[8] was for a European MSS system, but the apogee could be phased to occur at any longitude so that cities in high latitudes, but arbitrary longitude, could operate to an MSS satellite in Molniya orbit.

SIDEBAR

The Van Allen radiation belts were named after the discoverer of these belts, James A. Van Allen, who was the principal investigator for a radiation measurement package on the United States' first successful artificial earth satellite, *Explorer 1*, launched in January 1958.

The radiation belts consist of high-intensity protons and electrons that are temporarily trapped in the earth's magnetic field. While the trapped electrons can have energies up to 7 MeV (seven million electron volts), the trapped protons can have energies up to 500 MeV[10].

more critical if the bit flip is permanent, that is, "latch-up" occurs in a set position from which it cannot be changed.

The relative motion between the liquid core and the solid mantle and outer crust above it generates the earth's magnetic field. The magnetic field lines stretch out around the earth as shown schematically in Figure 10.7. While generally symmetrical close to the earth, the magnetic field lines of the earth become distorted further out from the earth due to interaction with the energy flowing toward the earth from the sun. The boundary where the solar atmosphere and the earth's magnetic field meet far out in space is called the bow shock, much like the pressure waves concentrating in front of the wing of an aircraft. Since the earth's magnetic and geographic poles are not coincident, the magnetic equator (and magnetic latitudes) will be different from the geographic equator (and geographic latitudes). The geomagnetic latitude ϕ can be computed from[11]:

$$\phi = \arcsin[\sin\alpha \sin 78.5° + \cos\alpha \cos 78.5° \cos(69° + \beta)] \tag{10.5}$$

where α is geographic latitude and β is geographic longitude. North and east coordinates are considered positive, and south and west coordinates negative.

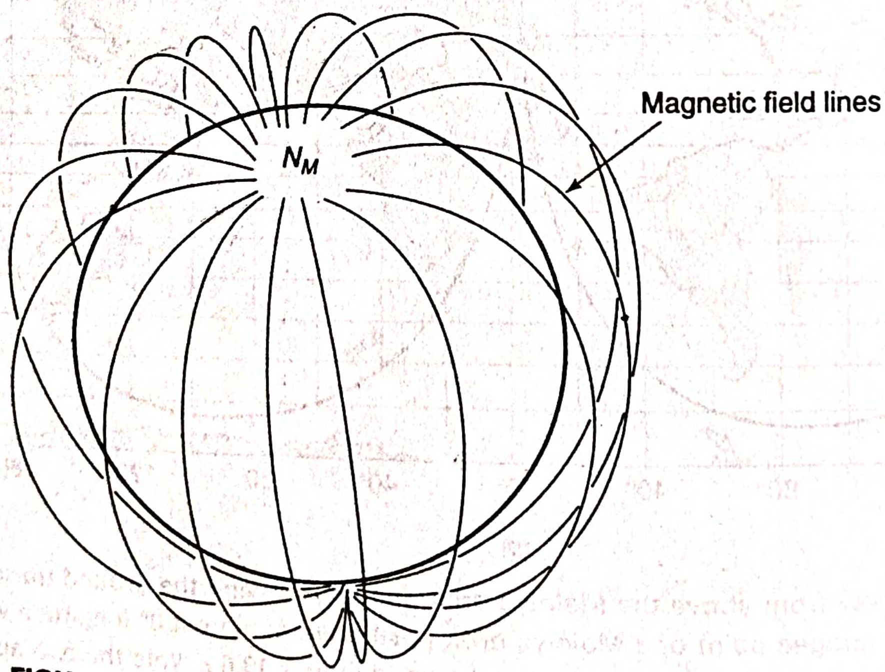


FIGURE 10.7 Representation of the magnetic field lines that flow between the north and south magnetic poles of the earth. The earth has a strong magnetic field due to having a liquid core that is spinning at a different rate than the solidified outer shell. The magnetic poles, however, are not coincident with the geomagnetic poles and so the magnetic equator is not located in the same position as the geographical equator. Sometimes the geomagnetic latitudes are referred to as *dip latitudes* since they will correspond to the dip in the magnetic field at that point. N_M is the north magnetic pole.

The electrons and protons become ensnared in the earth's magnetic field when their kinetic energy cannot overcome the trapping effect of magnetic lines of force at the given point of encounter in space. Since the magnetic field strength decreases with increase in altitude on a given radial from the center of the earth, only the electrons are trapped in the higher reaches of the earth's environs (>10,000 km altitude above the earth) since the field forces are relatively low at these altitudes. Both electrons and the higher energy protons are trapped lower down in the earth's atmosphere ~200 to 10,000 km[10], where the field is relatively more intense. The radiation levels induced by the electrons and protons fluctuate wildly with latitude, longitude, altitude, and with the sunspot cycle.

SIDEBAR

Sunspots are disturbances on the surface of the sun. Sunspots appear to generate huge outflows of energy from the sun and the amount of energy closely follows the number of sunspots—or rather groups of sunspots—which can be counted on the surface of the sun. The sunspot count, and hence the level of energy, varies with a mean period of about 11 years, although the actual cycle spans a 22-year *Hale cycle* as the magnetic field lines associated with the sunspot activity on the sun's surface reverse every 11 years. The 11-year sunspot cycle period is *not* constant. The period has been as short as 9.5 years and as long as 12.5 years[12]. The first cycle that has been given an official number is the 1755–1766 period. The last full solar cycle of the twentieth century (1986.8–1996.4) was labeled Cycle 22. A schematic of this cycle, showing the large variation in sunspot count that exists, is given in Figure 10.8. The turn of the twentieth century saw us in cycle 23 with the two-to-four year period of peak activity starting in the fall 1998 equinoctial period.

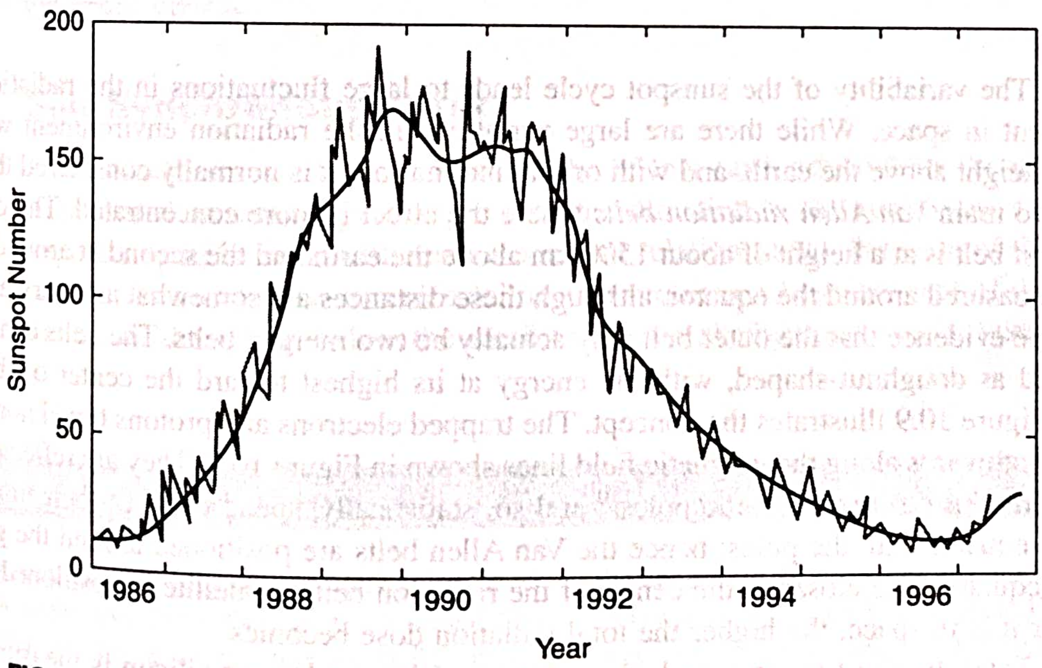


FIGURE 10.8 The general variation of the sunspot number over solar cycle 22. The smoothed sunspot number is averaged over several months. The fluctuations in the actual sunspot number are shown about this smoothed average. Not only does the sunspot count vary widely from month to month, it does so also from day to day. The higher the average sunspot number is, the larger the variation in actual sunspot number count is in general. Note the more rapid rise than decline in the average sunspot number count and the fairly long period when the sunspot activity was very high. Because of the "flat" nature of the sunspot maximum period (up to 4 years) it is usual to determine the sunspot periods from their minima.

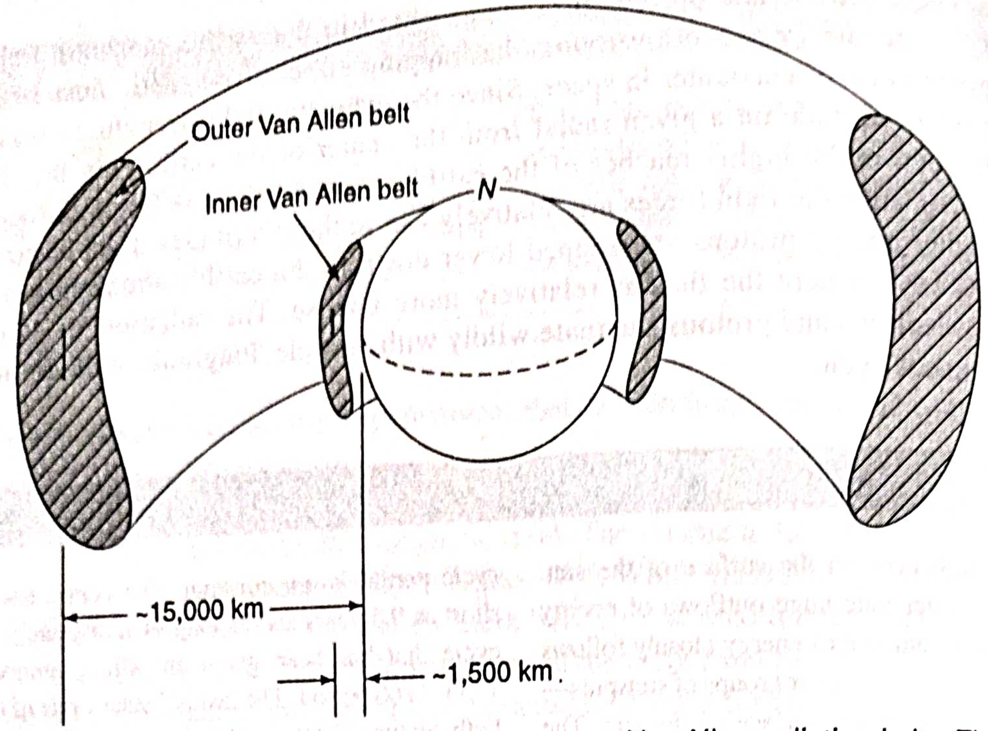


FIGURE 10.9 Pictorial representation of the two Van Allen radiation belts. The above schematic is a vertical "slice" through the radiation belts that exist around the equator. The shaded areas are the regions in the two belts where the radiation is at a maximum. The two principal NGSO regions lie "under" the first Van Allen radiation belt—the low earth orbit or LEO region—and "between" the two radiation belts—the medium earth orbit or MEO region—so as to avoid the highest radiation doses. However, radiation never falls to zero and exists in all areas (see reference 10).

The variability of the sunspot cycle leads to large fluctuations in the radiation environment in space. While there are large variations in the radiation environment with latitude, height above the earth, and with orbital inclination, it is normally considered that there are two main *Van Allen radiation belts* where the effect is more concentrated. The center of the first belt is at a height of about 1500 km above the earth and the second at around 15,000 km, measured around the equator, although these distances are somewhat arbitrary and there is some evidence that the outer belt may actually be two merged belts. The belts can be considered as doughnut-shaped, with the energy at its highest toward the center of the given belt. Figure 10.9 illustrates the concept. The trapped electrons and protons travel northwards and southwards along the magnetic field lines shown in Figure 10.7. They are reflected when they are close to the magnetic poles[10] and so, statistically, spend more of their time closer to the equator than the poles; hence the Van Allen belts are positioned around the geomagnetic equator. The closer to the center of the radiation belts a satellite is positioned and the longer it is in space, the higher the total radiation dose becomes.

Total dosage for semiconductors that are fabricated using silicon is measured with a unit called the krad(Si). A rad(Si) is a unit of energy absorbed by silicon from radiation and it is equivalent to 0.01 J/kg[10]. Radiation in near-earth space is highly variable. It changes both with height above the earth and with the inclination of the orbit with respect to the equatorial plane. Since the radiation is concentrated at the equator, satellites that are in equatorial orbits will receive a higher dosage than those that are in polar orbits will. In a like manner, as the orbital height moves from very close to the earth (300 km) outward for the first few thousand kilometers, the radiation dose will increase. Table 10.2 gives some typical examples of total radiation dosage for a LEO satellite designed for a 10-year operational lifetime.

TABLE 10.2 Typical Total Doses for Various Orbits

	Orbital height (km)		
Orbital type (degrees)	800	1100	2000
Polar orbit (90°)	30 krad(Si)	100 krad(Si)	>500 krad(Si)
Equatorial orbit (0°)			>2000 krad(Si)

The data are based on a 10-year mission life using silicon-based electronics in a satellite with a 2.5 mm thick aluminum skin.

Source: Data extracted from the text of reference 10.

Choosing an orbit that has a reduced level of radiation can therefore reduce the potential for radiation damage. Where this is not possible, then either radiation hardened (*rad-hard*) devices must be selected for the satellite or suitable shielding employed. Both are expensive options, the former because of the fabrication costs and the latter because radiation shields can be heavy and are nonproductive elements of the payload. Developing electronic devices that can withstand total radiation doses of 1 Mrad(Si) is possible with rad-hard technologies but newer techniques for approaching these levels of radiation hardening of devices will be needed for constellations of dozens of satellites. New, relatively cheap production processes have been shown to provide consistent shielding to 100 krad(Si) total dosage[10] and it is likely that such techniques, plus local site shielding with aluminum strips, will be largely employed for the foreseeable future. The same approach is being used for protection against single-event upsets.

Sun Synchronous Orbit

A sun synchronous orbit is a special form of low earth orbit where the plane of the orbit maintains a constant aspect angle with the direction to the sun. Some satellite missions require a specific orbit with such a constant relation to the direction of the sunlight. One example is an earth resources satellite that requires a large amount of direct sunlight to illuminate the region below the satellite so that photographs can be taken. This satellite

SIDEBAR

With the ever-smaller integrated circuits being developed for flight operations, there is an increased likelihood that the linear energy transfer (LET) that is generated by the heavy ion collision will cause a single-event ... thresholds of 37 MeV-cm^2/mg and the incidence of heavy ions with a LET exceeding 37 MeV-cm^2/mg is very rare[10]. The potential for latch-up will also reduce as technology advances permit devices to be used that

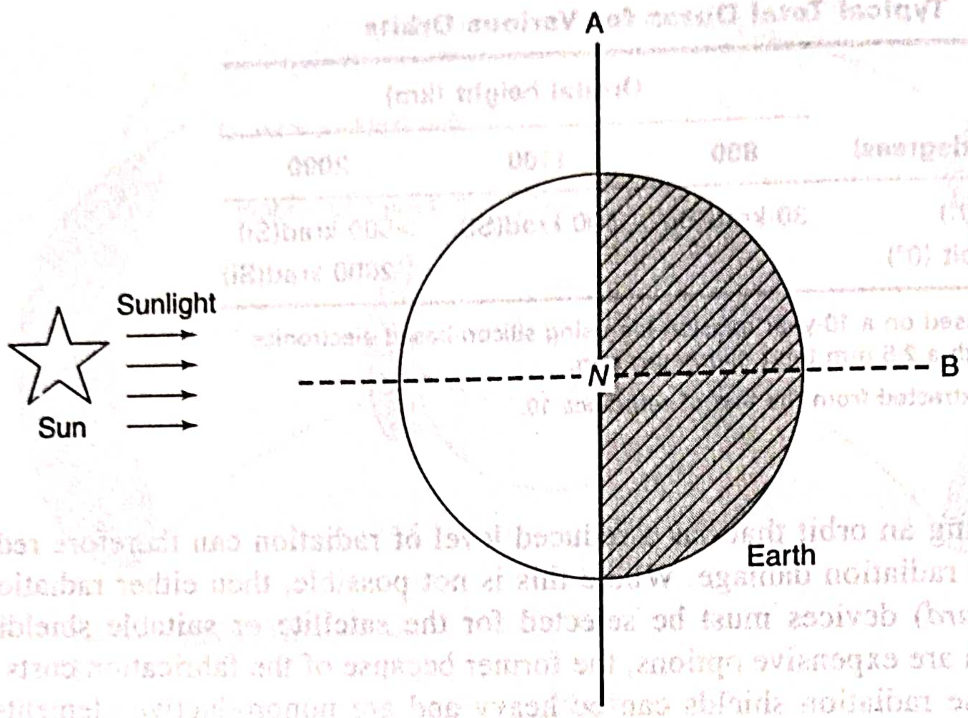


FIGURE 10.10 Examples of two sun synchronous orbits. In the illustration above, the earth is viewed from above the North Pole, *N*, with the sunlight illuminating the left side of the earth. Two sun synchronous orbits are shown. Orbit B is designed so that it will always have one-half of the orbit with the sun almost directly behind it; orbit A is designed to be always within sight of the sun—the so-called sunset–sunrise orbit since it will be orbiting the terminator. The terminator is the line that divides night from day.

would be in orbit B in Figure 10.10. Another example of a satellite needing this same orbit is a meteorological satellite, where images of the clouds and their directions of motion are critical in developing forecasts. While communications satellites have returned the greatest tangible investment returns for their owners, it is arguable that meteorological satellites have led directly to huge savings in human life, as well as to less property damage and farm animal destruction, in extreme weather situations. The early meteorological satellites (e.g., TIROS) were in LEO sun synchronous orbits, but all recent meteorological satellites are in GEO orbits to provide more instantaneous and continuous coverage. Other satellites that employ sun synchronous orbits are surveillance satellites.

Some surveillance satellites use orbit B of Figure 10.10, so that the maximum illumination is provided once per orbit. Others use orbit A of Figure 10.10. This particular "sunset–sunrise" orbit always has the satellite illuminated by the sun while the region below it has the sun at almost grazing incidence. There are two advantages in this orbit. First, the satellite need not have a large battery capacity for eclipse operations since it is always illuminated. Second, since the shadows are so long in the region being surveyed, changes in terrain or structures will be immediately obvious.

SIDEBAR

Before synthetic aperture radars were orbited, the sunset–sunrise orbit was used advantageously to detect changes in terrain following natural disasters such as earthquakes. The long shadow of a German V2 rocket allowed it to be detected by a reconnaissance aircraft for the first time at Peenemunde toward the end of the Second World War. Similarly, the ill-fated USSR moon rocket was detected on its launch pad by a surveillance satellite using the shadow it cast.

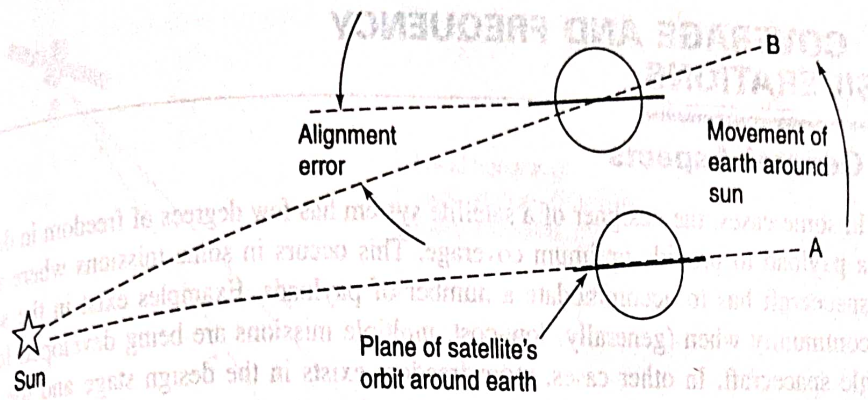


FIGURE 10.11 Illustration of alignment changes of the orbital plane of a satellite due to the movement of the earth around the sun. In the illustration above, a satellite has been launched into an LEO orbit (position A) in which the sun is directly overhead on the sunward side of the orbit. The view is from above the North Pole of the sun, orthogonal to the plane of the earth's orbit around the sun. When the earth has moved to position B, the plane of the satellite's orbit—which is fixed in inertial space—now has an alignment error with respect to the sun. If it is essential that the orbital plane of the satellite always be in line with the direction to the sun on a day-to-day basis over a long period, then the plane of the satellite's orbit will need to change at the same rate that the alignment error is increasing. That is, the plane of the orbit will need to "precess" to match the movement of the earth around the sun.

Figure 10.11 illustrates how the sun synchronous orbit is achieved. If a satellite is in a perfectly circular LEO orbit over the poles of the earth, a carefully timed launch would put the orbit in such a position that the sun is directly behind the satellite on the sunward side of the first orbit. This is position A of Figure 10.11. However, a short while later, the earth will have moved in its orbit around the sun and the plane of the satellite's orbit (now in position B) will no longer be aligned with the direction of the sunlight. In order to make the satellite's orbital plane always keep pace with the apparent change in position of the sun, it must be launched into a retrograde orbit. A retrograde orbit has a velocity component in a westerly direction. In practice, a LEO satellite launched into an orbit with an inclination of close to 98° to the equator (measured counter clockwise from the equator looking east) will move the orbital plane in time to the earth's movement around the sun. Elliptical orbits with different retrograde inclinations (see the constellation Ellipso in Section 10.5) will also yield sun synchronous orbits. The change (rotation) in the orbital plane is called *precession*. A key advantage of a sun synchronous orbit is that it will repeat its track every half day. It can therefore be used to make measurements at given times of the day and night so that correlation exercises can be attempted.

A sun synchronous orbit will pass over almost all of the earth at one time or another. Determining the instantaneous surface area of the planet seen by the satellite and over which information is required—or to which communications is to be established—is another issue. This portion of the earth's surface is called the coverage area or coverage region.

SIDEBAR

One example of a spacecraft in a sun synchronous orbit was the *Mars Explorer* spacecraft, which was put into a sun synchronous orbit around Mars in 1998. The orbit was used to measure temperature at 2 A.M. and 2 P.M. local time equivalents over the same region so that local heating effects and cooling effects could be accurately tracked.

10.4 DELAY AND THROUGHPUT CONSIDERATIONS

Delay in a communications link is not normally a problem unless the interactions between the users are very rapid—a few milliseconds apart in response time. Long delays, such as those associated with manned missions to the moon, required the development of agreed procedures, much like tactical military or police communications requires specific hand-off code words such as "over" to signal the end of one user's input. For most commercial satellite links that are over long distances, particularly those with satellites in geostationary orbit, the main problem was not delay, but echo. A mismatched transmission line will always have a reflected signal. If the mismatch is large, a strong echo will return. Over a GEO satellite link, the echo arrives back in the telephone headset about half a second after the speaker has spoken, and usually while the speaker is still speaking. This will interrupt the speaker and the conversation becomes fragmented. The development of echo suppressors and, even better, echo cancellers, solved the problem. Figure 10.25 illustrates the one-way propagation time for a typical LEO, MEO, and GEO system.

Based on the calculations shown in Figure 10.25, the time delay for a signal passing between LEO user 1 and LEO user 2 in the same instantaneous coverage is 5.4 ms (2.7 ms up and 2.7 ms down) and the go and return (round-trip) delay between the two users is twice this at 10.8 ms. It is rare, however, for a user to be immediately underneath a LEO satellite and, for LEO satellites in higher orbits, the round-trip delays due to propagation time can be more than double this. Globalstar, which has a maximum path length from the satellite to the user of 2500 km, will have a maximum round-trip delay time of 33 ms. For GEO users, the up and down (forward) link delay is typically 230 ms with the round-trip delay 460 ms. However, Figure 10.25 does not tell the whole story. Most MSS systems use voice compression to reduce the bandwidth required for a single voice channel. The coded bit rates for a single voice channel range from 2.4 kbit/s for Globalstar to 6.25 kbit/s for Iridium[13]. The *vocoders* sample the incoming analog voice signal and produce excellent, low data rate digital reproductions—but at a price in delay. The access scheme can also add additional delay. If the channel is operated in a simplex fashion, i.e., you cannot send at the same time as you are receiving, there can be a delay in response. The Iridium TDMA access mode uses a time division duplexing (TDD) scheme. A TDD scheme allows transmissions to occur for a certain period (while receive functions are off) and then transmissions cease while receive operations are in use. In the present Iridium TDMA access scheme, eight users share a frequency assignment and, within this frequency channel, share a 45-ms transmit frame and a 45-ms receive frame. There can therefore be up to 90 ms between transmission and reception of specific parts of a message. On the Iridium satellite, the onboard processing system translates the received signal to baseband, the header address information is read, and the appropriate route selected for onward transmission. The baseband signal is then reformatted, up-converted to the RF band, and transmitted. All of this takes time. The forward delay (ground-to-satellite plus satellite-to-ground) within the same instantaneous coverage averaged 153 ms in the initial operational tests of Iridium. A transoceanic link delay using intersatellite links averaged 253 ms—almost the same as for a GEO satellite link. Delay can also have an adverse effect on the throughput of the signal, as noted in Chapter 9. If the protocol used in the link is not adapted for the particular delay environment, appreciable reduction in throughput will occur. Customer acceptance of a service has been found to be driven by three prime factors: *access ability* (i.e., can the required connection be obtained immediately on request?), *availability* (i.e., once connected, will the call be dropped?), and *performance* (i.e., is the error rate low and the throughput high?). Pricing will attract customers but it will not keep them for long if all three prime factors are not met.

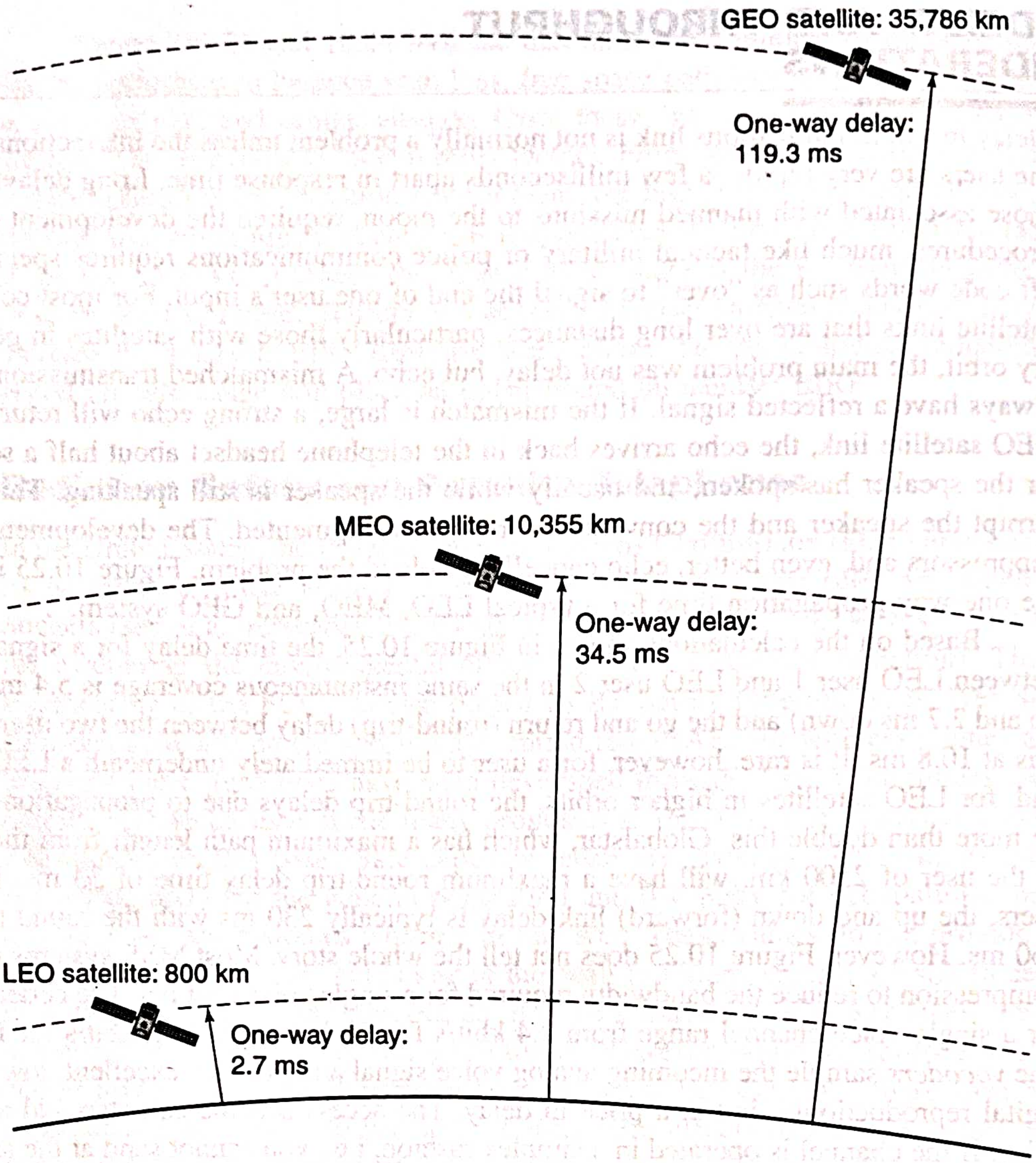


FIGURE 10.25 One-way propagation delay for the three orbits: LEO, MEO, and GEO. The one-way delay figures shown above have been calculated assuming the radio signal propagates at the speed of light in a vacuum, i.e., 3×10^8 m/s. That is, no account has been taken of any delay due to the refractive index of the atmosphere not being unity. Also, no account has been taken of any processing delay imposed on the signal from any source coding, channel coding, modulation, or access scheme used.

As a final element in the discussion on delay, it is worth noting the challenges that face system designers when intersatellite links (ISLs) are employed to relay signals around a LEO constellation. It is a fairly straightforward matter to design an ISL to connect two GEO satellites or a LEO satellite to a GEO satellite: the relative motions are not that large. Consider now a LEO system attempting to establish connections across the constellation. The connections will have to be both in plane (i.e., around the same orbit plane of that particular ring of satellites) and across planes. When the satellites are close to the equator, the orbital planes are at their furthest separation and the rate of change between two LEO satellites traveling in the same direction is at a minimum. As satellites move closer to the poles, the more rapidly they have to steer their ISL antennas to maintain contact. In some operational modes, Iridium switches off the across-plane ISL links when the spacecraft are above latitudes of about 60°[13]. In no case, however, can Iridium maintain an ISL link between planes where the satellites are moving

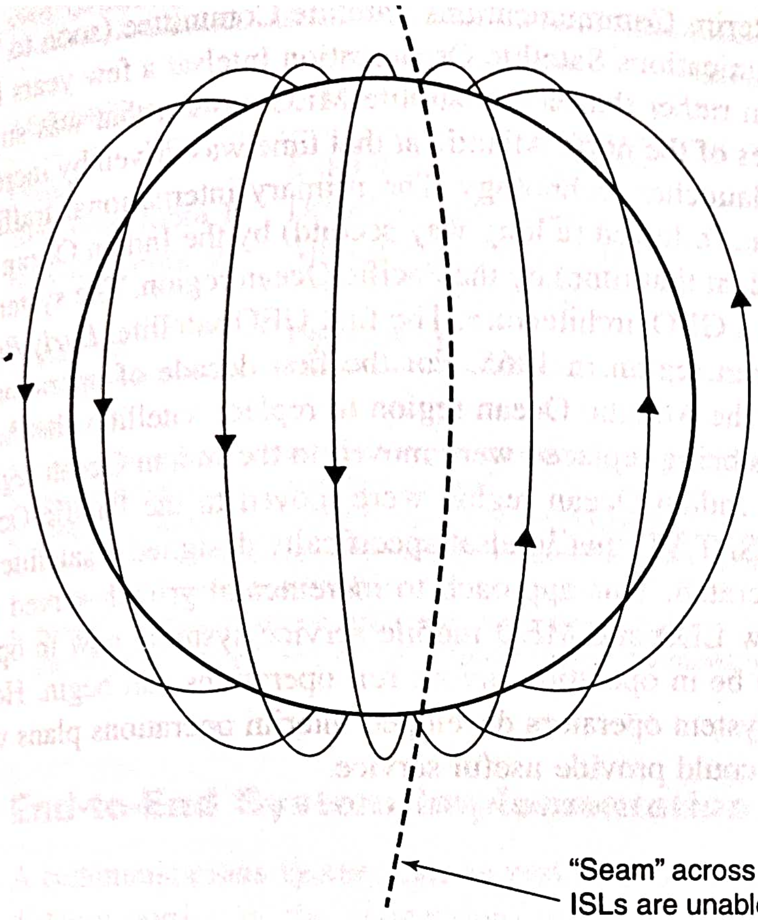


FIGURE 10.26 Schematic of the ISL seam in the Iridium constellation. The Iridium satellites are in an orbit that is close to polar (86.5° inclination). There are four ISL antennas on each satellite that are used to communicate with adjacent satellites. The ISLs operate at 23 GHz and use solid reflector tracking antennas. The inertial mass of the antennas combined with the need to have a stable satellite platform for the normal communications mode to the earth limits the rate of change of the tracking mechanism. Satellites across the seam are traveling at a closing speed of about 36,000 mph (~58,000 km/h) and it is likely that only lightweight optical ISLs will be able to track at the angular rates of change required across an LEO seam.

in opposite directions. There will therefore be a *seam* in the constellation across which no ISL links can operate. This is illustrated in Figure 10.26. The revised Teledesic system (once 840 satellites, then 288 satellites, and now fewer than 200 satellites) has reportedly been designed to operate across the LEO "seam" and so it is likely that the ISLs will be optical and not microwave. Optical ISL antennas are much smaller and lighter than microwave ISL antennas and so impose less tracking restrictions due to inertial forces when under acceleration. Whether or not to use ISLs; whether to design to operate across the seam if ISLs are used; selecting an orbital height, number of satellites visible at any instant, coverage region; etc.; all interact in the overall system design. We will now look at other system considerations that can affect the design of the satellite network in other respects.

10.5 SYSTEM CONSIDERATIONS

There are four important factors that influence the design of any satellite communication system: incremental growth, interim operations, (satellite) replenishment options, and end-to-end system implementation.

Incremental Growth

The 1964 decision by the Interim Communications Satellite Committee (soon to become The International Telecommunications Satellite Organization Intelsat a few years later) to select a GEO satellite system rather than a 12-satellite MEO system that was supported by major entities on both sides of the north Atlantic at that time was driven by incremental growth plans as well as by launcher technology. The primary international traffic route was across the Atlantic Ocean, followed (a long way second) by the Indian Ocean region, and (an even longer way third, at that time) by the Pacific Ocean region. The system could be grown incrementally with a GEO architecture. The first GEO satellite, *Early Bird*, was placed over the Atlantic Ocean region in 1965. For the first decade of operations, new satellites were launched into the Atlantic Ocean region to replace satellites that had been operating there. The satellites being replaced were moved to the Indian Ocean region and the satellites replaced in the Indian Ocean region were moved to the Pacific Ocean region. It was not until INTELSAT VII that Intelsat specifically designed a satellite for the Pacific Ocean region from scratch. This approach to incremental growth served Intelsat well. By comparison, the new LEO and MEO mobile service systems now in operation require all of the satellites to be in operation before full operations can begin. However, most of the LEO and MEO system operators developed interim operations plans where a reduced number of satellites could provide useful service.

Interim Operations

Interim operations for LEO and MEO systems serve two functions: they can bring a service on line gradually, introducing the technology to the market while teething problems are sorted out; and they can act as fall back plans should multiple satellite failures occur over a short period. Nearly all of the LEO and MEO systems undertook such interim operations. Orbcomm began commercial operations with less than half of its 36-satellite constellation in place, thus becoming the first commercial LEO system to establish a revenue stream. Globalstar began with 32 out of the planned 48-satellite constellation and New ICO plans to start operations with six out of the planned ten-satellite constellation. Iridium, since it uses ISLs to complete the network, required all 66 satellites to be available before beginning beta testing in November 1998. The technical planning for interim operations includes relaxing the number of satellites visible to any user at any particular time, which lowers the number of satellites required to complete the constellation. The elevation angle minimum for users is also usually lowered, the gaps between operational satellites in the same plane are made symmetrical, and the orbits adjusted if possible to maximize coverage over those parts of the day when user service requests are highest. Most LEO constellations have at least four satellites per plane and multiple spacecraft launches are used in the constellation buildup. A Pegasus launch vehicle carries eight Orbcomm satellites into orbit, A Delta II carries five Globalstar satellites, and a Proton carries seven Iridium satellites into LEO. When a satellite fails in service, there is an in-orbit spare to take its place. If more than one satellite fails in a plane, additional satellites must be launched to replenish the system.

Replenishment Options

Launching five or more satellites to replace one failed satellite makes little economic sense. As a result, the LEO service providers use smaller rockets to replenish their system. Table 10.5 lists the primary and replenishment launchers used by the Big LEO systems and Orbcomm.

TABLE 10.5 Primary and Replenishment Launchers for the Big-LEO Systems and Orbcomm

LEO/MEO system	Primary launchers (number per launch)	Secondary/replenishment launchers (number per launch)
Iridium	Delta II (5) and Proton (7)	Long March (2)
Globalstar [note 1]	Delta II (5) and Soyuz (4)	[Note 2]
New ICO [note 3]	Atlas IIAS (1), Proton (1), Zenit (1), Delta III (1)	[Note 4]
Orbcomm	Pegasus (8)	Taurus (2)

Notes:

1. Globalstar initially selected Zenit but canceled the launch services contract when the first launch failed with 12 satellites aboard.
2. Globalstar has not selected a replenishment launcher.
3. The Zenit rockets for New ICO are to be launched from a floating platform. Sealaunch is a joint venture of Boeing and Hughes.
4. New ICO satellites are so large, and the orbital altitude so high, that only one satellite is launched per rocket. Any of the selected rockets could act as replenishment launch vehicles.

End-to-End System Implementation

A communications system can be part of a larger network (e.g., just providing the long-distance portion of the connection) or it can provide the full end-to-end system implementation, from user to user. AT&T and Intelsat, when they were first set up, did not provide end-to-end service: AT&T provided long-distance capacity for local telephone companies and Intelsat provided satellite capacity for entities such as AT&T to carry their international traffic. Neither company interacted directly with the end user. Indeed, specific laws or protocols prevented this from happening.

The design of an NGSO system will be heavily influenced by the decision on whether or not to provide service directly to the end user. It will also be impacted by the decision on whether or not to include established telephone companies in the delivery of the service. By their very nature, mobile satellite systems have committed to serve the end user directly. However, different approaches have been taken with regard to including established telephone companies. Two examples of organizations that took opposite decisions are Globalstar and Iridium. Globalstar elected not to bypass existing telephone companies while Iridium did. These decisions led to a very different architecture for the two systems, which will be discussed in the next section.

10.6 OPERATIONAL NGSO CONSTELLATION DESIGNS

Seven satellite constellation designs are reviewed briefly in the following discussion, four MSS offerings with multiple beams, one with single beam coverage providing both two-way services and one-way store-and-forward services, and two Internet-multimedia satellite systems.

Ellipso

The Ellipso constellation drew from studies of the world's population distribution and the potential market for MSS users. Figure 10.27 (abstracted from data in reference 19)

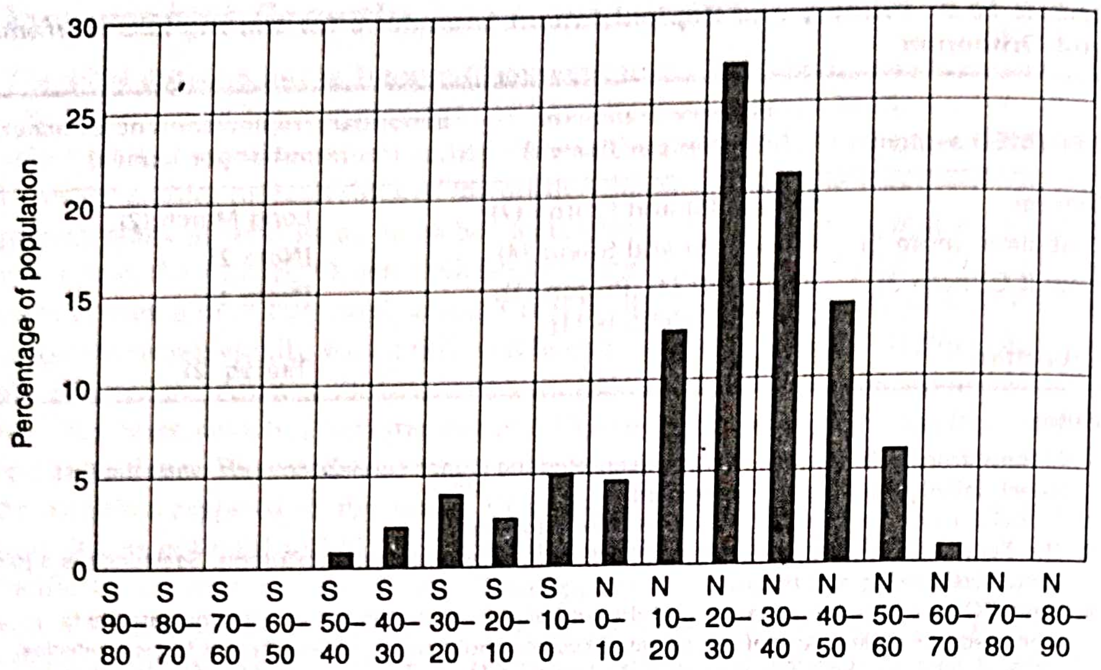


FIGURE 10.27 Percentage of the world's population living in the given latitude ranges (data extracted from reference 19 and reproduced with permission). The data in the figure show that more than 85% of the world's population lives in the northern hemisphere. Designing a satellite system that spends most of its time in the northern hemisphere would therefore cover the world's population more efficiently.

shows that more than 85% of the world's population lives north of the equator. Additional studies[19] concluded that an equatorial constellation of MEO satellites could serve the bulk of the world's population. Ellipso therefore adopted an incremental approach to their service offering. The first set of satellites would be in a circular equatorial orbit. The second set would be in elliptical equatorial orbit, with the ellipticity of the orbit designed to provide dwell times over the regions of greater demand. The third set of satellites would be in sun synchronous 3-hour orbits inclined at 116.6° to provide coverage over the highly industrialized northern hemisphere regions. The equatorial orbit groups of the Ellipso system are called Concordia™ and the sun synchronous group is called Borealis™. Details can be found in Table 10.6. The Ellipso spacecraft is based on the Boeing GPS satellite bus and up to five satellites can be launched by a single rocket. No onboard processing is performed; the signals received at the satellite are transponded down to gateway earth stations for onward routing via the terrestrial PSTN or satellite network. No ISLs are used.

Globalstar

In a similar manner to Ellipso, Globalstar elected to develop a constellation that was aimed at the populous regions of the earth. The Globalstar orbital planes are therefore inclined at 52° to the equator, thus ignoring the sparsely populated high-latitude regions. To minimize the power requirements of the user handset, the constellation altitude was lowered to just below the first Van Allen radiation belt. This increased the total number of satellites needed to 48. No onboard processing or ISLs are used; the signals received at the satellite are simply transponded down and the gateway earth stations process the signals for the onward routing (see Figure 10.28). Like Ellipso, service over water is

TABLE 10.6 System Parameters of Five NGSO Constellations Aimed at Data and Voice Communications

System parameter	Ellipso	Globalstar	New ICO	Iridium	Orbcomm
Number of planes	1→3→5	6	2	6	4→5
Satellites per plane	1 × 7 then 1 × 7 and 2 × 3 then 1 × 7, 2 × 3, 2 × 5	8	5	11	4 × 8 then 4 × 8 and 1 × 4
Total complement	23	48	10	66	36
Orbital inclination	3 at 0°, 2 at 116.6°	52°	45°	86.5°	4 at 45°, 1 at 72°
Orbit type	1 circular (0°), 2 elliptical (0°), 2 sun synchronous	Circular	Circular	Circular	Circular (45° and 72°)
Orbital height (km)	1 circular 8050, 2 elliptical 6149–8050, 2 sun synchronous 633–7605	1414	10,255	780	775
Spot beams per satellite	61	16	163	48	1
Satellite lifetime	5 to 7 years	~7.5 years	~12 years	5 to 7 years	5 to 7 years

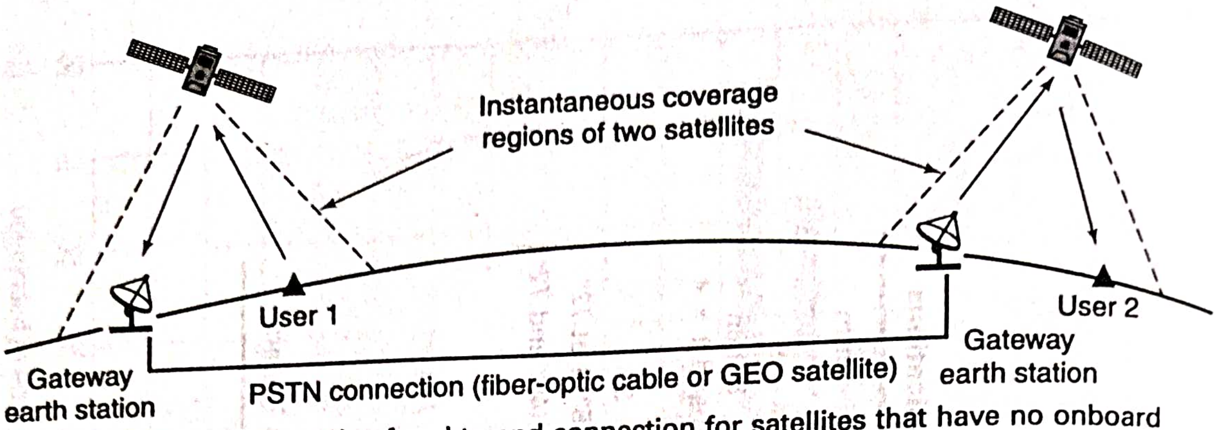


FIGURE 10.28 Schematic of end-to-end connection for satellites that have no onboard processing or ISLs. User 1 is in a different instantaneous coverage region than that occupied by user 2. The signal from user 1 is picked up by the gateway earth station and relayed to the gateway earth station of user 2. The signal is then sent up to the satellite from the second gateway earth station and then down to user 2. If user 1 or user 2 is using a fixed telephone (or computer) the signal would simply pass over the regular PSTN circuits and not via the space segment. Because the users must be in line-of-sight contact with a gateway earth station, no maritime traffic can be picked unless the ship is close to land (and a gateway earth station).

restricted to coastal regions where the satellite is within radio range of a gateway earth station.

New ICO

ICO Global is the company that was spun off from the International Maritime Satellite Organization (Inmarsat); New ICO is the company that emerged from bankruptcy protection in 2000. Inmarsat was initially set up solely for the purpose of providing reliable communications to maritime traffic. Later, Inmarsat also provided aeronautical services, in addition to priority links for safety communications, whether on land or sea. New ICO, although primarily aimed at the LMS market (Land Mobile Services), also needed to provide capacity for maritime links. New ICO elected not to include ISLs in their system architecture nor any significant onboard processing. Since a LEO constellation would not provide maritime coverage without ISLs, a higher orbit was necessary. If little onboard processing is used, traffic routing from mobile to mobile would have to be carried out at the gateway earth stations (as it is for Ellipso and Globalstar) necessitating a double-hop link. A double-hop link involves two uplinks and two downlinks. (A double hop is used in Figure 10.28; two different earth–space links are used to complete the connection.) A double-hop configuration is not feasible for a GEO constellation since the overall delay would be completely unacceptable at about 1 s. New ICO therefore adopted a MEO constellation. An inclination of 45° is used, but since the orbit altitude is so high, full global coverage is possible. Toward the end of 2000, it was learned that New ICO was modifying the payload and service offerings to provide two-way Internet-like connections in a joint venture with Teledesic. At the time of going to press, it is unclear how this synergy will evolve.

Iridium

The genesis of Iridium was formed around the need to communicate from anywhere to anywhere on the surface of the world, even where no telecommunications infrastructure existed. The system therefore must be stand-alone. From this—and the need for a low power handset—came the concept of first 77, and then 66, almost-polar orbiting LEO satellites linked via ISLs. Each of the satellites in the constellation acts as a switching node. Uplink signals are received and demodulated at the satellite using onboard processing to recover individual data packets at baseband so that the header information can be read. Using this information and links to the network control stations, the next node for each packet is determined and the packet is reformatted with the next address. The baseband data packet is then processed and up-converted for transmission either to the ground at L band directly to another Iridium user or at 20 GHz to a gateway earth station, or over one of the four ISL links (at 23 GHz) to the next satellite in the chain. Onboard processing is needed to carry out the entire message routing and formatting functions.

Orbcomm

Many research organizations and businesses need to obtain data from locations that are either inaccessible on a regular basis or are moving within areas without good cellular telephone coverage. Examples are buoys measuring water characteristics in rivers and at sea, and delivery trucks. Tracking of high value cargo on trucks is another application that needs to send a short message to a central station at regular intervals. A GPS receiver on the cargo determines its location and this information is sent with an ID number via an Orbcomm satellite. If the truck carrying the cargo is hijacked, its route can be followed and the truck intercepted.

Much of this information is neither required in real time nor does it need a high capacity link. Orbcomm developed their system around this requirement and have orbited a constellation of satellites with both two-way data communications and store-and-forward capabilities (see Table 10.6). The satellites are lightweight (40 kg) and simple in design and execution[3]. A single beam is used to develop the instantaneous coverage and no onboard processing is used.

A terminal that is within the coverage area of a satellite and a gateway station (which includes almost all of the United States) can send short messages to the gateway station in real time. The message length is limited to a few hundred bytes. A terminal that has data waiting to be uploaded for store-and-forward listens for the passage of a satellite and then uploads its data when the satellite is in view. The data, in the form of a packet with the address of the intended recipient, are stored and transmitted to a gateway station for onward transmission to the recipient when the satellite is within range of the gateway station. Orbcomm satellites carry short messages, with a relatively high cost per transmitted bit. The system is therefore most attractive to users who want to send a small number of high value bits, such as requests for help in emergency situations or tracking information for high value cargo.

None of the five NGSO constellations above were initially designed to carry traffic at rates higher than 10 kbps. This is not adequate for Internet access, which has emerged as a potentially important requirement in mobile systems. Two NGSO constellations that addressed this market from the outset are Skybridge and Teledesic.

Skybridge

Skybridge evolved a similar approach to coverage as Globalstar, by selecting an inclined orbit that covers the major population densities. Like Globalstar, Skybridge satellites carry a nonprocessing payload and do not have intersatellite links; so all traffic is transponded down to the gateway earth stations for processing and onward routing. However, Skybridge satellites are intended to carry wideband traffic and therefore use frequencies above 10 GHz. They chose to employ the same Ku-band frequencies as the FSS service in GEO uses: 12.75–14.5 GHz on the uplink and 10.7–12.75 GHz on the downlink. To allow successful coordination with existing FSS GEO systems, they elected to prevent any operations (up or down) whenever a satellite look angle is within 10° of the GEO orbital plane. This requirement led to a relatively large number of satellites (80 vs 48) for the constellation. The decision not to use ISLs also required a very large number of gateway earth stations (on the order of 200). Skybridge also uses the concept of a fixed earth cell (see Figure 10.29). More details of Skybridge can be found in Table 10.7.

Teledesic

Teledesic started from the same precept as Iridium, but is designed for Internet-like data traffic rather than voice communication. Any user can access any other user or ISP (Internet service provider) independent of location and the existing telecommunications

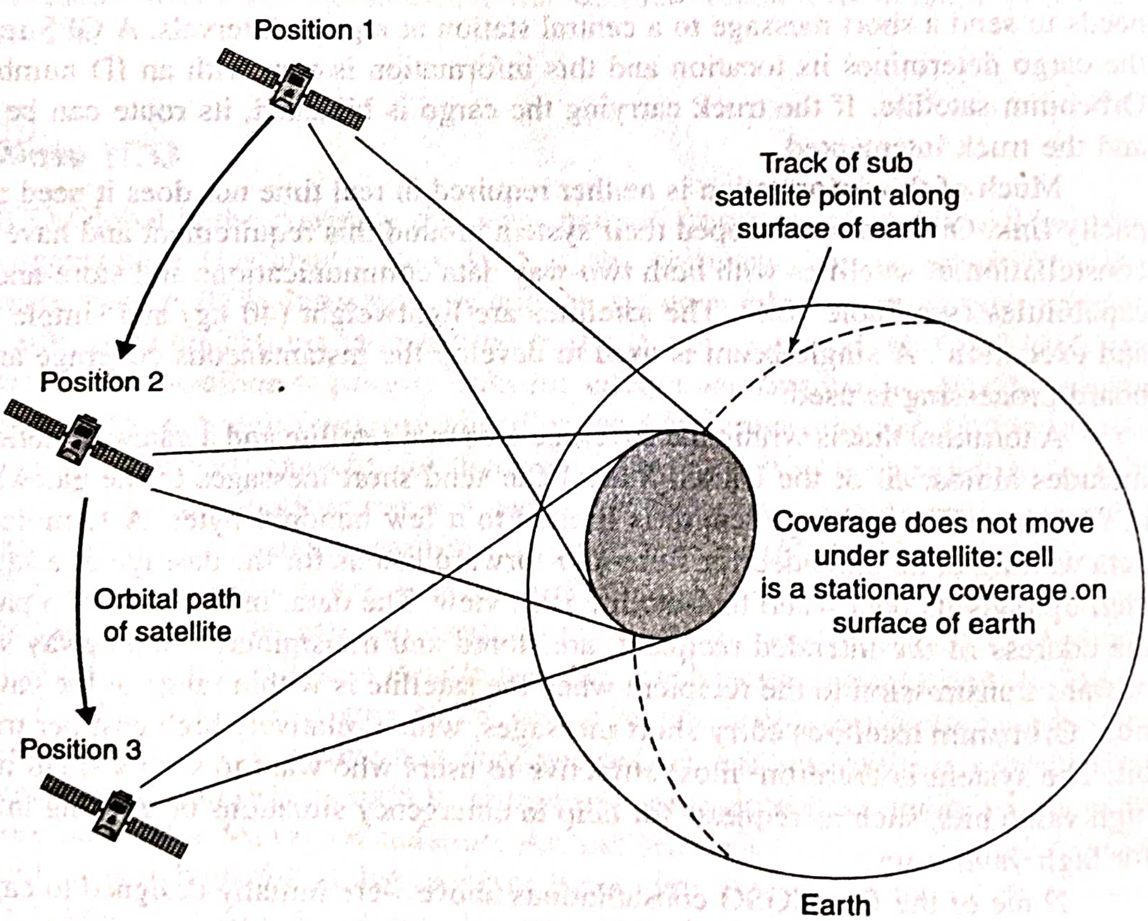


FIGURE 10.29 Concept of a stationary cell. Unlike the coverage of the NGSO satellite show in Figure 10.14, a stationary coverage (or "fixed earth cell") of an NGSO satellite does not move with the satellite. The phased array antenna on the satellite steers the beam, while the satellite transits, to keep the coverage on the surface of the earth constant. As the satellite moves between positions 1, 2, and 3, stationary coverage is maintained on the surface of the earth. Separate antennas are used for communications coverage and gateway links. In this way, a gateway need not necessarily be within a given stationary coverage.

TABLE 10.7 System Parameters of Two NGSO Constellations Aimed at Internet Multimedia Communications

System parameter	Skybridge	Teledesic
Number of planes	20	12
Satellites per plane	4	24
Total complement	80	288
Orbital inclination	53°	~90°
Orbit type	circular	circular
Orbital height (km)	1469	~1400
Spot-beams per satellite	18	—
Satellite lifetime	~7 years	~7 years

infrastructure. The concept of Teledesic is to provide a complete worldwide data communications system above the surface of the earth using satellites, instead of on the earth's surface using fiber-optic cables. This requirement dictated the use of wideband data links, onboard processing, and ISL links. To avoid the necessity of coordinating with existing systems, Teledesic chose to move their operations completely into Ka band. As noted earlier, to reduce the impact of rain, Teledesic also limited the elevation angle at which users could access the satellites (the mask angle) to 40°. The initial Teledesic constellation had a complement of 840 satellites (22 planes with 40 operational satellites per plane) plus 40 spare satellites in orbit. The orbital altitude was later moved up from 700 km to about 1400 km, which reduced the number of planes to 12, with 24 operational satellites in each plane (see Table 10.7). The early estimates of Teledesic's system cost were between \$9 B and \$12 B, using 840 satellites. Reduction in the number of satellites to 288 lowered the cost significantly, and further reductions in the number of satellites seem likely to make the cost of creating the system more acceptable.

Other companies are seeking to provide Internet access from satellites with lower cost solutions than those of Teledesic and Skybridge. One company aiming to do so for a "mere" \$2.6 B is Virtual Geosatellite[20]. However, for a given bit rate, no proposed system is lower in cost than a GEO alternative[21]. The geostationary earth orbit has the unique characteristic of providing data transfer by satellite at the lowest cost per bit. None of the currently scheduled or operating LEO and MEO satellite constellations has been able to demonstrate a significant added value from the use of their particular service when a commercial return on investment is required. There is a clear military requirement for many of the new constellations—from anywhere to anywhere—without any intervening infrastructure, but the growth in terrestrial cellular systems and optical fiber links has removed much of the potential commercial demand for these new services. At the turn of the twenty-first century, more than 90% of all Internet traffic flowed through about 30 metropolitan areas. If these conurbations are connected via optical fibers or through high-powered spot beam antennas from GEO, the remaining traffic is what a LEO or MEO system would pick up. The same is true for cellular telephony: what the major cities do not provide leaves very little traffic for a high priced LEO or MEO alternative. Table 10.8 gives some early 2000 data on Internet traffic centers.