CHAPTER 1

Introduction

Elliott D. Kaplan
The MITRE Corporation

1.1 Introduction

Navigation is defined as the science of getting a craft or person from one place to another. Each of us conducts some form of navigation in our daily lives. Driving to work or walking to a store requires that we employ fundamental navigation skills. For most of us, these skills require utilizing our eyes, common sense, and landmarks. However, in some cases where a more accurate knowledge of our position, intended course, or transit time to a desired destination is required, navigation aids other than landmarks are used. These may be in the form of a simple clock to determine the velocity over a known distance or the odometer in our car to keep track of the distance traveled. Some other navigation aids transmit electronic signals and therefore are more complex. These are referred to as *radionavigation aids*.

Signals from one or more radionavigation aids enable a person (herein referred to as the *user*) to compute their position. (Some radionavigation aids provide the capability for velocity determination and time dissemination as well.) It is important to note that it is the user's radionavigation receiver that processes these signals and computes the position fix. The receiver performs the necessary computations (e.g., range, bearing, and estimated time of arrival) for the user to navigate to a desired location. In some applications, the receiver may only partially process the received signals, with the navigation computations performed at another location.

Various types of radionavigation aids exist, and for the purposes of this text they are categorized as either ground-based or space-based. For the most part, the accuracy of ground-based radionavigation aids is proportional to their operating frequency. Highly accurate systems generally transmit at relatively short wavelengths, and the user must remain within line of sight (LOS), whereas systems broadcasting at lower frequencies (longer wavelengths) are not limited to LOS but are less accurate. Early spaced-based systems (namely, the U.S. Navy Navigation Satellite System—referred to as Transit—and the Russian Tsikada system)[1] provided a two-dimensional high-accuracy positioning service. However, the frequency of obtaining a position fix is dependent on the user's latitude. Theoretically,

1. Transit was decommissioned on December 31, 1996, by the U.S. government. At the time of this writing, Tsikada was still operational.

a Transit user at the equator could obtain a position fix on the average of once every 110 minutes, whereas at 80° latitude the fix rate would improve to an average of once every 30 minutes [1]. Limitations applicable to both systems are that each position fix requires approximately 10 to 15 minutes of receiver processing and an estimate of the user's position. These attributes were suitable for shipboard navigation because of the low velocities, but not for aircraft and high-dynamic users [2]. It was these shortcomings that led to the development of the U.S. Global Positioning System (GPS).

1.2 Condensed GPS Program History

In the early 1960s, several U.S. government organizations, including the Department of Defense (DOD), the National Aeronautics and Space Administration (NASA), and the Department of Transportation (DOT), were interested in developing satellite systems for three-dimensional position determination. The optimum system was viewed as having the following attributes: global coverage, continuous/all weather operation, ability to serve high-dynamic platforms, and high accuracy. When Transit became operational in 1964, it was widely accepted for use on low-dynamic platforms. However, due to its inherent limitations (cited in the preceding paragraphs), the Navy sought to enhance Transit or develop another satellite navigation system with the desired capabilities mentioned earlier. Several variants of the original Transit system were proposed by its developers at the Johns Hopkins University Applied Physics Laboratory. Concurrently, the Naval Research Laboratory (NRL) was conducting experiments with highly stable space-based clocks to achieve precise time transfer. This program was denoted as Timation. Modifications were made to Timation satellites to provide a ranging capability for two-dimensional position determination. Timation employed a sidetone modulation for satellite-to-user ranging [3–5].

At the same time as the Transit enhancements were being considered and the Timation efforts were underway, the Air Force conceptualized a satellite positioning system denoted as System 621B. It was envisioned that System 621B satellites would be in elliptical orbits at inclination angles of 0°, 30°, and 60°. Numerous variations of the number of satellites (15–20) and their orbital configurations were examined. The use of pseudorandom noise (PRN) modulation for ranging with digital signals was proposed. System 621B was to provide three-dimensional coverage and continuous worldwide service. The concept and operational techniques were verified at the Yuma Proving Grounds using an inverted range in which pseudosatellites or *pseudolites* (i.e., ground-based satellites) transmitted satellite signals for aircraft positioning [3–6]. Furthermore, the Army at Ft. Monmouth, New Jersey, was investigating many candidate techniques, including ranging, angle determination, and the use of Doppler measurements. From the results of the Army investigations, it was recommended that ranging using PRN modulation be implemented [5].

In 1969, the Office of the Secretary of Defense (OSD) established the Defense Navigation Satellite System (DNSS) program to consolidate the independent development efforts of each military service to form a single joint-use system. The OSD also established the Navigation Satellite Executive Steering Group, which was charged with determining the viability of the DNSS and planning its development. From this effort, the system concept for NAVSTAR GPS was formed. The NAVSTAR GPS program was developed by the GPS Joint Program Office (JPO) in El Segundo, California [5]. At the time of this writing, the GPS JPO continued to oversee the development and production of new satellites, ground control equipment, and the majority of U.S. military user receivers. Also, the system is now most commonly referred to as simply *GPS*.

1.3 GPS Overview

Presently, GPS is fully operational and meets the criteria established in the 1960s for an optimum positioning system. The system provides accurate, continuous, worldwide, three-dimensional position and velocity information to users with the appropriate receiving equipment. GPS also disseminates a form of Coordinated Universal Time (UTC). The satellite constellation nominally consists of 24 satellites arranged in 6 orbital planes with 4 satellites per plane. A worldwide ground control/monitoring network monitors the health and status of the satellites. This network also uploads navigation and other data to the satellites. GPS can provide service to an unlimited number of users since the user receivers operate passively (i.e., receive only). The system utilizes the concept of one-way time of arrival (TOA) ranging. Satellite transmissions are referenced to highly accurate atomic frequency standards onboard the satellites, which are in synchronism with a GPS time base. The satellites broadcast ranging codes and navigation data on two frequencies using a technique called code division multiple access (CDMA); that is, there are only two frequencies in use by the system, called L1 (1,575.42 MHz) and L2 (1,227.6 MHz). Each satellite transmits on these frequencies, but with different ranging codes than those employed by other satellites. These codes were selected because they have low cross-correlation properties with respect to one another. Each satellite generates a short code referred to as the coarse/acquisition or C/A code and a long code denoted as the precision or P(Y) code. (Additional signals are forthcoming. Satellite signal characteristics are discussed in Chapter 4.) The navigation data provides the means for the receiver to determine the location of the satellite at the time of signal transmission, whereas the ranging code enables the user's receiver to determine the transit (i.e., propagation) time of the signal and thereby determine the satellite-to-user range. This technique requires that the user receiver also contain a clock. Utilizing this technique to measure the receiver's three-dimensional location requires that TOA ranging measurements be made to four satellites. If the receiver clock were synchronized with the satellite clocks, only three range measurements would be required. However, a crystal clock is usually employed in navigation receivers to minimize the cost, complexity, and size of the receiver. Thus, four measurements are required to determine user latitude, longitude, height, and receiver clock offset from internal system time. If either system time or height is accurately known, less than four satellites are required. Chapter 2 provides elaboration on TOA ranging as well as user position, velocity, and time (PVT) determination.

GPS is a dual-use system. That is, it provides separate services for civil and military users. These are called the Standard Positioning Service (SPS) and the Precise

Positioning Service (PPS). The SPS is designated for the civil community, whereas the PPS is intended for U.S. authorized military and select government agency users. Access to the GPS PPS is controlled through cryptography. Initial operating capability (IOC) for GPS was attained in December 1993, when a combination of 24 prototype and production satellites was available and position determination/timing services complied with the associated specified predictable accuracies. GPS reached full operational capability (FOC) in early 1995, when the entire 24 production satellite constellation was in place and extensive testing of the ground control segment and its interactions with the constellation was completed. Descriptions of the SPS and PPS services are presented in the following sections.

1.3.1 PPS

The PPS is specified to provide a predictable accuracy of at least 22m (2 drms, 95%) in the horizontal plane and 27.7m (95%) in the vertical plane. The distance root mean square (drms) is a common measure used in navigation. Twice the drms value, or 2 drms, is the radius of a circle that contains at least 95% of all possible fixes that can be obtained with a system (in this case, the PPS) at any one place. The PPS provides a UTC time transfer accuracy within 200 ns (95%) referenced to the time kept at the U.S. Naval Observatory (USNO) and is denoted as UTC (USNO) [7, 8]. Velocity measurement accuracy is specified as 0.2 m/s (95%) [4]. PPS measured performance is addressed in Section 7.7.

As stated earlier, the PPS is primarily intended for military and select government agency users. Civilian use is permitted, but only with special U.S. DOD approval. Access to the aforementioned PPS position accuracies is controlled through two cryptographic features denoted as antispoofing (AS) and selective availability (SA). AS is a mechanism intended to defeat deception jamming through encryption of the military signals. Deception jamming is a technique in which an adversary would replicate one or more of the satellite ranging codes, navigation data signal(s), and carrier frequency Doppler effects with the intent of deceiving a victim receiver. SA had intentionally degraded SPS user accuracy by *dithering* the satellite's clock, thereby corrupting TOA measurement accuracy. Furthermore, SA could have introduced errors into the broadcast navigation data parameters [9]. SA was discontinued on May 1, 2000, and per current U.S. government policy is to remain off. When it was activated, PPS users removed SA effects through cryptography [4].

1.3.2 SPS

The SPS is available to all users worldwide free of direct charges. There are no restrictions on SPS usage. This service is specified to provide accuracies of better than 13m (95%) in the horizontal plane and 22m (95%) in the vertical plane (global average; signal-in-space errors only). UTC (USNO) time dissemination accuracy is specified to be better than 40 ns (95%) [10]. SPS measured performance is typically much better than specification (see Section 7.7).

At the time of this writing, the SPS was the predominant satellite navigation service in use by millions throughout the world.

1.4 GPS Modernization Program

In January 1999, the U.S. government announced a new GPS modernization initiative that called for the addition of two civil signals to be added to new GPS satellites [11]. These signals are denoted as L2C and L5. The L2C signal will be available for nonsafety of life applications at the L2 frequency; the L5 signal resides in an aeronautical radionavigation service (ARNS) band at 1,176.45 MHz. L5 is intended for safety-of-life use applications. These additional signals will provide SPS users the ability to correct for ionospheric delays by making dual frequency measurements, thereby significantly increasing civil user accuracy. By using the carrier phase of all three signals (L1 C/A, L2C, and L5) and differential processing techniques, very high user accuracy (on the order of millimeters) can be rapidly obtained. (Ionospheric delay and associated compensation techniques are described in Chapter 7, while differential processing is discussed in Chapter 8.) The additional signals also increase the receiver's robustness to interference. If one signal experiences high interference, then the receiver can switch to another signal. It is the intent of the U.S. government that these new signals will aid civil, commercial, and scientific users worldwide. One example is that the combined use of L1 (which also resides in an ARNS band) and L5 will greatly enhance civil aviation.

During the mid to late 1990s, a new military signal called M code was developed for the PPS. This signal will be transmitted on both L1 and L2 and is spectrally separated from the GPS civil signals in those bands. The spectral separation permits the use of noninterfering higher power M code modes that increase resistance to interference. Furthermore, M code will provide robust acquisition, increased accuracy, and increased security over the legacy P(Y) code.

Chapter 4 contains descriptions of the legacy (C/A code and P(Y) code) and modernized signals mentioned earlier.

At the time of this writing, it was anticipated that both M code and L2C will be on orbit when the first Block IIR-M ("R" for replenishment, "M" for modernized) satellite is scheduled to be launched. (The Block IIR-M will also broadcast all legacy signals.) The Block IIF ("F" for follow on) satellite is scheduled for launch in 2007 and will generate all signals, including L5. Figure 1.1 provides an overview of GPS signal evolution. Figures 1.2 and 1.3 depict the Block IIR-M and Block IIF satellites, respectively.

At the time of this writing, the GPS III program was underway. This program was conceived in 2000 to reassess the entire GPS architecture and determine the necessary architecture to meet civil and military user needs through 2030. It is envisioned that GPS III will provide submeter position accuracy, greater timing accuracy, a system integrity solution, a high data capacity intersatellite crosslink capability, and higher signal power to meet military antijam requirements. At the time of this writing, the first GPS III satellite launch was planned for U.S. government fiscal year 2013.

CHAPTER 2

Fundamentals of Satellite Navigation

Elliott D. Kaplan and Joseph L. Leva
The MITRE Corporation

Dennis Milbert
NOAA (retired)

Mike S. Pavloff
Raytheon Company

2.1 Concept of Ranging Using TOA Measurements

GPS utilizes the concept of TOA ranging to determine user position. This concept entails measuring the time it takes for a signal transmitted by an emitter (e.g., foghorn, radiobeacon, or satellite) at a known location to reach a user receiver.

This time interval, referred to as the signal propagation time, is then multiplied by the speed of the signal (e.g., speed of sound or speed of light) to obtain the emitter-to-receiver distance. By measuring the propagation time of the signal broadcast from multiple emitters (i.e., navigation aids) at known locations, the receiver can determine its position. An example of two-dimensional positioning is provided next.

2.1.1 Two-Dimensional Position Determination

Consider the case of a mariner at sea determining his or her vessel's position from a foghorn. (This introductory example was originally presented in [1] and is contained herein because it provides an excellent overview of TOA position determination concepts.) Assume that the vessel is equipped with an accurate clock and the mariner has an approximate knowledge of the vessel's position. Also, assume that the foghorn whistle is sounded precisely on the minute mark and that the vessel's clock is synchronized to the foghorn clock. The mariner notes the elapsed time from the minute mark until the foghorn whistle is heard. The foghorn whistle propagation time is the time it took for the foghorn whistle to leave the foghorn and travel to the mariner's ear. This propagation time multiplied by the speed of sound (approximately 335 m/s) is the distance from the foghorn to the mariner. If the foghorn signal took 5 seconds to reach the mariner's ear, then the distance to the foghorn is 1,675m. Let this distance be denoted as $R1$. Thus, with only one measurement, the mariner knows that the vessel is somewhere on a circle with radius $R1$ centered about the foghorn, which is denoted as Foghorn 1 in Figure 2.1.

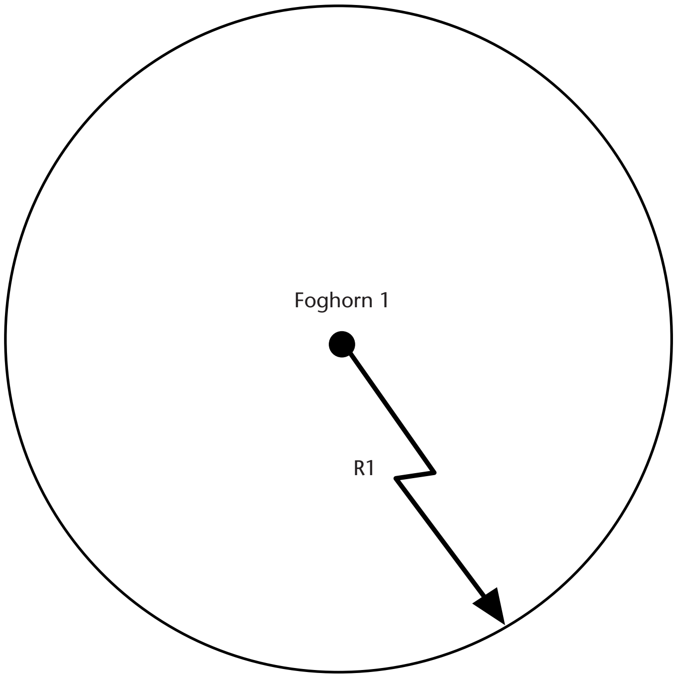


Figure 2.1 Range determination from a single source. (*After:* [1].)

Hypothetically, if the mariner simultaneously measured the range from a second foghorn in the same way, the vessel would be at range *R*1 from Foghorn 1 and range *R*2 from Foghorn 2, as shown in Figure 2.2. It is assumed that the foghorn transmissions are synchronized to a common time base and the mariner has knowledge of both foghorn whistle transmission times. Therefore, the vessel relative to the foghorns is at one of the intersections of the range circles. Since it was assumed that the mariner has approximate knowledge of the vessel's position, the unlikely fix can be discarded. Resolving the ambiguity can also be achieved by making a range measurement to a third foghorn, as shown in Figure 2.3.

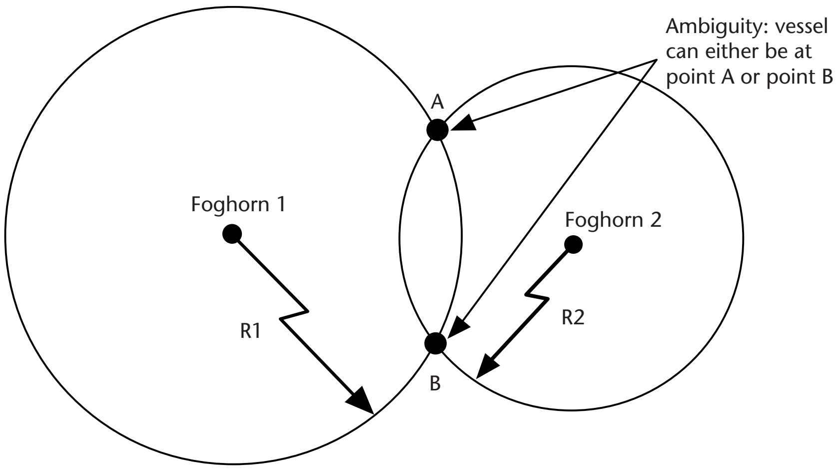


Figure 2.2 Ambiguity resulting from measurements to two sources. (*After:* [1].)

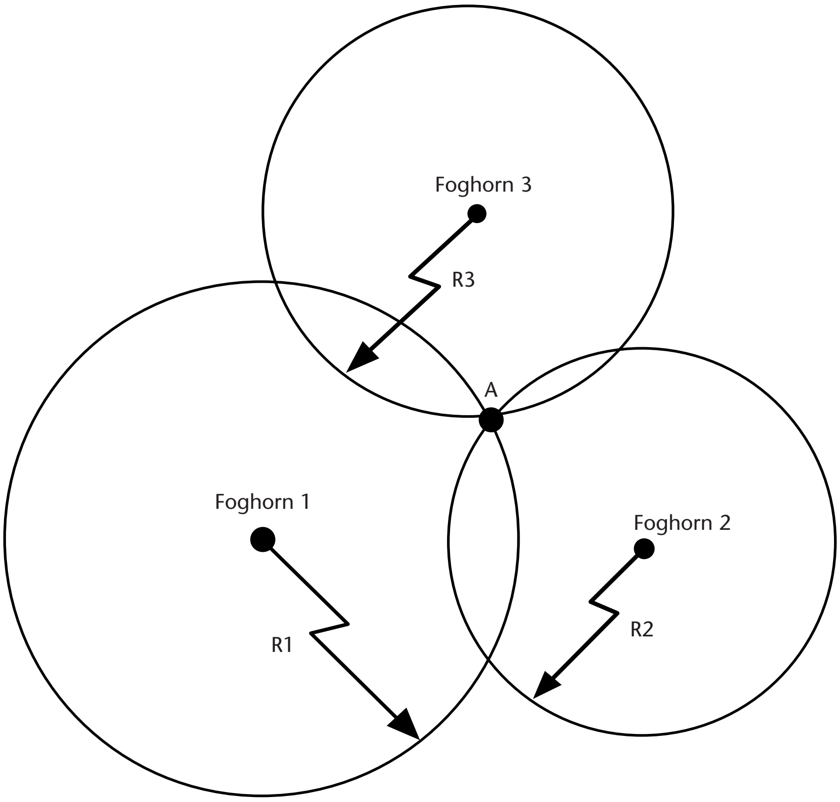


Figure 2.3 Position ambiguity removal by additional measurement. (*After:* [1].)

2.1.1.1 Common Clock Offset and Compensation

This development assumed that the vessel's clock was precisely synchronized with the foghorn time base. However, this might not be the case. Let us presume that the vessel's clock is advanced with respect to the foghorn time base by 1 second. That is, the vessel's clock believes the minute mark is occurring 1 second earlier. The propagation intervals measured by the mariner will be larger by 1 second due to the offset. The timing offsets are the same for each measurement (i.e., the offsets are common) because the same incorrect time base is being used for each measurement. The timing offset equates to a range error of 335m and is denoted as ε in Figure 2.4. The separation of intersections C, D, and E from the true vessel position, A, is a function of the vessel's clock offset. If the offset could be removed or compensated for, the range circles would then intersect at point A.

2.1.1.2 Effect of Independent Measurement Errors on Position Certainty

If this hypothetical scenario were realized, the TOA measurements would not be perfect due to errors from atmospheric effects, foghorn clock offset from the foghorn time base, and interfering sounds. Unlike the vessel's clock offset condition cited earlier, these errors would be generally independent and not common to all measurements. They would affect each measurement in a unique manner and result in inaccurate distance computations. Figure 2.5 shows the effect of independent

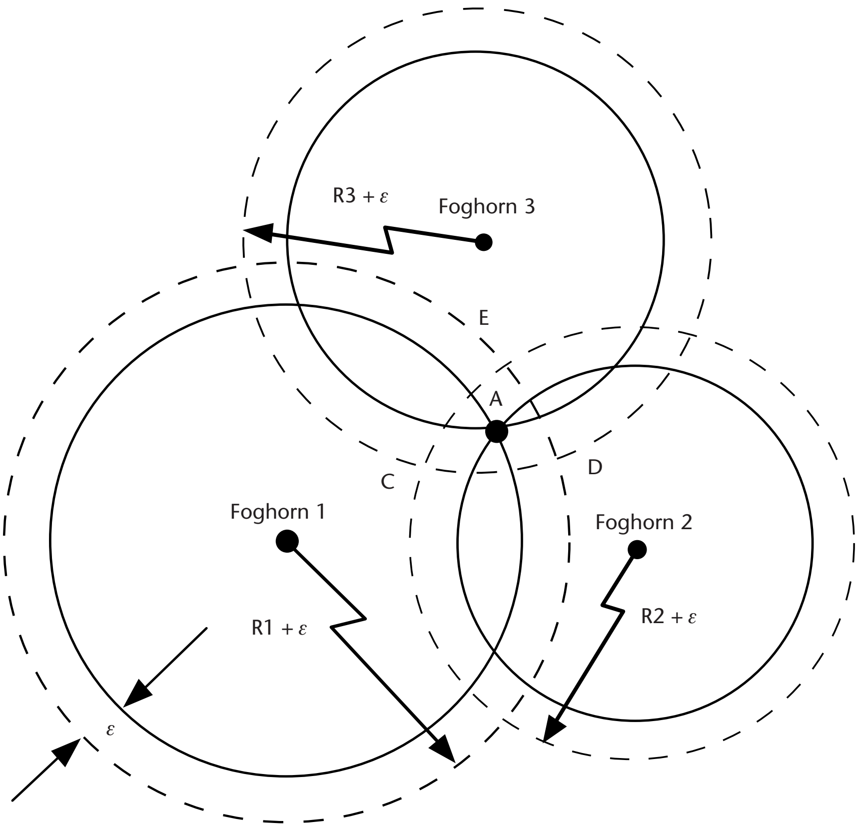


Figure 2.4 Effect of receiver clock offset on TOA measurements. (*After:* [1].)

errors (i.e., ε_1, ε_2, and ε_3) on position determination assuming foghorn timebase/mariner clock synchronization. Instead of the three range circles intersecting at a single point, the vessel location is somewhere within the triangular error space.

2.1.2 Principle of Position Determination Via Satellite-Generated Ranging Signals

GPS employs TOA ranging for user position determination. By making TOA measurements to multiple satellites, three-dimensional positioning is achieved. We will observe that this technique is analogous to the preceding foghorn example; however, satellite ranging signals travel at the speed of light, which is approximately 3 10^8 m/s. It is assumed that the satellite ephemerides are accurate (i.e., the satellite locations are precisely known).

2.1.2.1 Three-Dimensional Position Location Via Intersection of Multiple Spheres

Assume that there is a single satellite transmitting a ranging signal. A clock onboard the satellite controls the timing of the ranging signal broadcast. This clock and others onboard each of the satellites within the constellation are effectively synchronized to an internal system time scale denoted as GPS system time (herein referred to as system time). The user's receiver also contains a clock that (for the moment) we assume

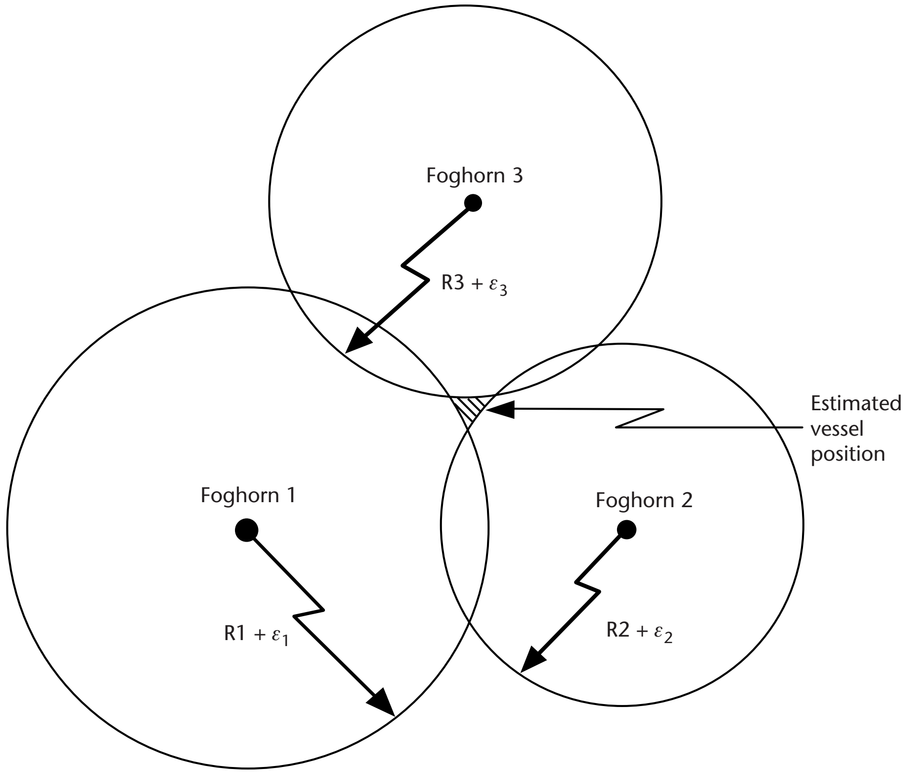
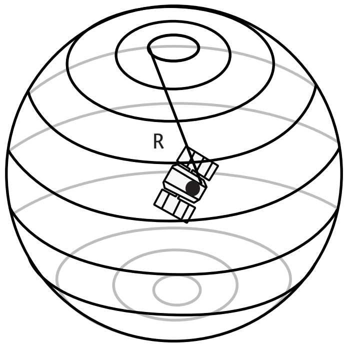


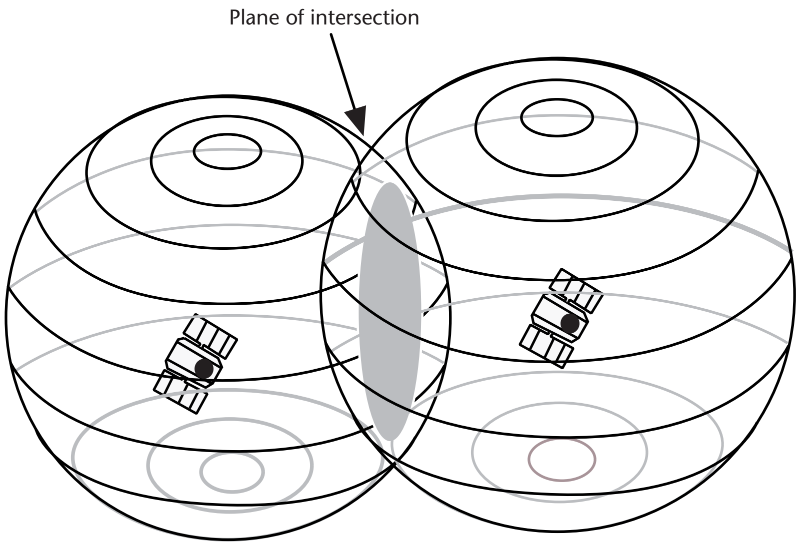
Figure 2.5 Effect of independent measurement errors on position certainty.

to be synchronized to system time. Timing information is embedded within the satellite ranging signal that enables the receiver to calculate when the signal left the satellite based on the satellite clock time. This is discussed in more detail in Section 2.4.1. By noting the time when the signal was received, the satellite-to-user propagation time can be computed. The product of the satellite-to-user propagation time and the speed of light yields the satellite-to-user range, R. As a result of this measurement process, the user would be located somewhere on the surface of a sphere centered about the satellite, as shown in Figure 2.6(a). If a measurement were simultaneously made using the ranging signal of a second satellite, the user would also be located on the surface of a second sphere that is concentric about the second satellite. Thus, the user would then be somewhere on the surface of both spheres, which could be either on the perimeter of the shaded circle in Figure 2.6(b) that denotes the plane of intersection of these spheres or at a single point tangent to both spheres (i.e., where the spheres just touch). This latter case could only occur if the user were collinear with the satellites, which is not the typical case. The plane of intersection is perpendicular to a line connecting the satellites, as shown in Figure 2.6(c).

Repeating the measurement process using a third satellite, the user is at the intersection of the perimeter of the circle and the surface of the third sphere. This third sphere intersects the shaded circle perimeter at two points; however, only one of the points is the correct user position, as shown in Figure 2.6(d). A view of the intersection is shown in Figure 2.6(e). It can be observed that the candidate locations are mirror images of one another with respect to the plane of the satellites. For



(a)



(b)

Figure 2.6 (a) User located on surface of sphere. (b) User located on perimeter of shaded circle. (*Source:* [2]. Reprinted with permission.) (c) Plane of intersection. (d) User located at one of two points on shaded circle. (*Source:* [2]. Reprinted with permission.) (e) User located at one of two points on circle perimeter.

a user on the Earth's surface, it is apparent that the lower point will be the true position. However, users that are above the Earth's surface may employ measurements from satellites at negative elevation angles. This complicates the determination of an unambiguous solution. Airborne/spaceborne receiver solutions may be above or below the plane containing the satellites, and it may not be clear which point to select unless the user has ancillary information.

2.2 Reference Coordinate Systems

To formulate the mathematics of the satellite navigation problem, it is necessary to choose a reference coordinate system in which the states of both the satellite and the

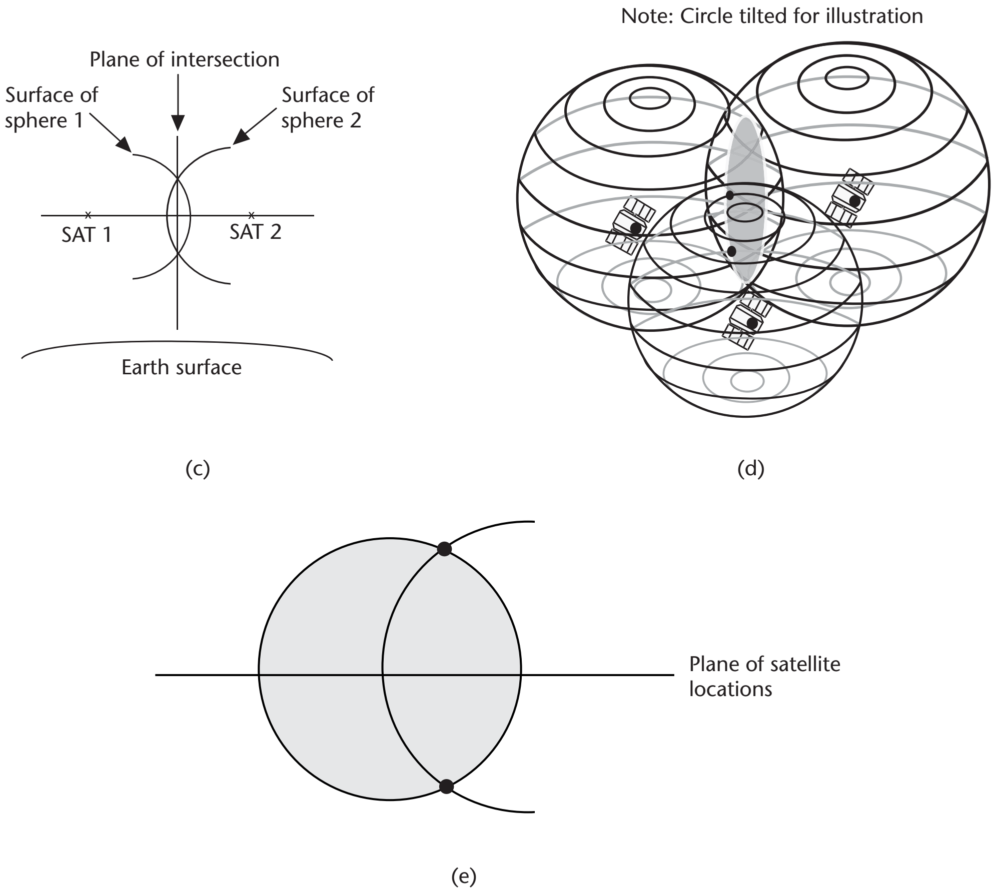


Figure 2.6 (continued.)

receiver can be represented. In this formulation, it is typical to describe satellite and receiver states in terms of position and velocity vectors measured in a Cartesian coordinate system. Two principal Cartesian coordinate systems are inertial and rotating systems. In this section, an overview is provided of the coordinate systems used for GPS.

2.2.1 Earth-Centered Inertial Coordinate System

For the purposes of measuring and determining the orbits of the GPS satellites, it is convenient to use an Earth-centered inertial (ECI) coordinate system, in which the origin is at the center of the mass of the Earth and whose axes are pointing in fixed directions with respect to the stars. A GPS satellite obeys Newton's laws of motion and gravitation in an ECI coordinate system. In typical ECI coordinate systems, the xy-plane is taken to coincide with the Earth's equatorial plane, the $+x$-axis is permanently fixed in a particular direction relative to the celestial sphere, the $+z$-axis is taken normal to the xy-plane in the direction of the north pole, and the $+y$-axis is chosen so as to form a right-handed coordinate system. Determination and subsequent prediction of the GPS satellite orbits are carried out in an ECI coordinate system.

One subtlety in the definition of an ECI coordinate system arises due to irregularities in the Earth's motion. The Earth's shape is oblate, and due largely to the gravitational pull of the Sun and the Moon on the Earth's equatorial bulge, the equatorial plane moves with respect to the celestial sphere. Because the x-axis is defined relative to the celestial sphere and the z-axis is defined relative to the equatorial plane, the irregularities in the Earth's motion would cause the ECI frame as defined earlier not to be truly inertial. The solution to this problem is to define the orientation of the axes at a particular instant in time, or *epoch*. The GPS ECI coordinate system uses the orientation of the equatorial plane at 1200 hours UTC (USNO) on January 1, 2000, denoted as the J2000 system. The $+x$-axis is taken to point from the center of the mass of the Earth to the direction of vernal equinox, and the y- and z-axes are defined as described previously, all at the aforementioned epoch. Since the orientation of the axes remains fixed, the ECI coordinate system defined in this way can be considered inertial for GPS purposes.

2.2.2 Earth-Centered Earth-Fixed Coordinate System

For the purpose of computing the position of a GPS receiver, it is more convenient to use a coordinate system that rotates with the Earth, known as an Earth-centered Earth-fixed (ECEF) system. In such a coordinate system, it is easier to compute the latitude, longitude, and height parameters that the receiver displays. As with the ECI coordinate system, the ECEF coordinate system used for GPS has its xy-plane coincident with the Earth's equatorial plane. However, in the ECEF system, the $+x$-axis points in the direction of 0° longitude, and the $+y$-axis points in the direction of 90°E longitude. The x-, y-, and z-axes therefore rotate with the Earth and no longer describe fixed directions in inertial space. In this ECEF system, the z-axis is chosen to be normal to the equatorial plane in the direction of the geographical North Pole (i.e., where the lines of longitude meet in the northern hemisphere), thereby completing the right-handed coordinate system.

GPS orbit computation software includes the transformations between the ECI and the ECEF coordinate systems. Such transformations are accomplished by the application of rotation matrices to the satellite position and velocity vectors in the ECI coordinate system, as described, for example, in [3]. The broadcast orbit computation procedure described in [4] and in Section 2.3 generates satellite position and velocity in the ECEF frame. Precise orbits from numerous computation centers also express GPS position and velocity in ECEF. Thus, with one exception, we may proceed to formulate the GPS navigation problem in the ECEF system without discussing the details of the orbit determination or the transformation to the ECEF system. This exception is consideration of the Sagnac effect on signal propagation in the rotating (noninertial) ECEF frame. (Section 7.2.3 contains an explanation of the Sagnac effect.)

As a result of the GPS navigation computation process, the Cartesian coordinates (x_u, y_u, z_u) of the user's receiver are computed in the ECEF system, as described in Section 2.4.2. It is typical to transform these Cartesian coordinates to latitude, longitude, and height of the receiver. In order to carry out this transformation, it is necessary to have a physical model describing the Earth.

2.2.3 World Geodetic System

The standard physical model of the Earth used for GPS applications is the DOD's World Geodetic System 1984 (WGS 84) [5]. One part of WGS 84 is a detailed model of the Earth's gravitational irregularities. Such information is necessary to derive accurate satellite ephemeris information; however, we are concerned here with estimating the latitude, longitude, and height of a GPS receiver. For this purpose, WGS 84 provides an ellipsoidal model of the Earth's shape, as shown in Figure 2.7. In this model, cross-sections of the Earth parallel to the equatorial plane are circular. The equatorial cross-section of the Earth has radius 6,378.137 km, which is the mean equatorial radius of the Earth. In the WGS 84 Earth model, cross-sections of the Earth normal to the equatorial plane are ellipsoidal. In an ellipsoidal cross-section containing the z-axis, the major axis coincides with the equatorial diameter of the Earth. Therefore, the semimajor axis, a, has the same value as the mean equatorial radius given previously. The minor axis of the ellipsoidal cross-section shown in Figure 2.7 corresponds to the polar diameter of the Earth, and the semiminor axis, b, in WGS 84 is taken to be 6,356.7523142 km. Thus, the eccentricity of the Earth ellipsoid, e, can be determined by

$$e = \sqrt{1 - \frac{b^2}{a^2}}$$

WGS 84 takes $e^2 = 0.00669437999014$. It should be noted that this figure is extremely close, but not identical, to the Geodetic Reference System 1980 (GRS 80) ellipsoid quantity of $e^2 = 0.00669438002290$. These two ellipsoids differ only by 0.1 mm in the semiminor axis, b.

Another parameter sometimes used to characterize the reference ellipsoid is the second eccentricity, e', which is defined as follows:

$$e' = \sqrt{\frac{a^2}{b^2} - 1} = \frac{a}{b}e$$

WGS 84 takes $e'^2 = 0.00673949674228$.

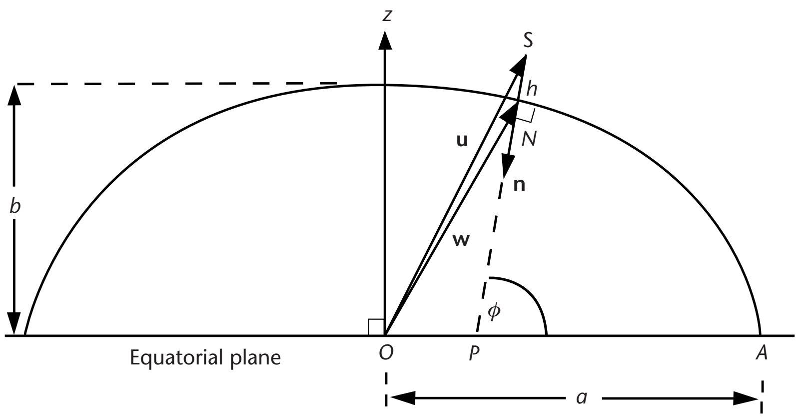


Figure 2.7 Ellipsoidal model of Earth (cross-section normal to equatorial plane).

Table 2.2 GPS Ephemeris Data Definitions

t_{0e}	Reference time of ephemeris
$\sqrt{a}$	Square root of semimajor axis
e	Eccentricity
i_0	Inclination angle (at time t_{0e})
Ω_0	Longitude of the ascending node (at weekly epoch)
ω	Argument of perigee (at time t_{0e})
M_0	Mean anomaly (at time t_{0e})
di/dt	Rate of change of inclination angle
$\dot{\Omega}$	Rate of change of longitude of the ascending node
Δn	Mean motion correction
C_{uc}	Amplitude of cosine correction to argument of latitude
C_{us}	Amplitude of sine correction to argument of latitude
C_{rc}	Amplitude of cosine correction to orbital radius
C_{rs}	Amplitude of sine correction to orbital radius
C_{ic}	Amplitude of cosine correction to inclination angle
C_{is}	Amplitude of sine correction to inclination angle

Table 2.3 Computation of a Satellite's ECEF Position Vector

(1)	$a = \left(\sqrt{a}\right)^2$	Semimajor axis
(2)	$n = \sqrt{\frac{\mu}{a^3}} + \Delta n$	Corrected mean motion, $\mu = 398{,}600.5 \quad 10^8\ \mathrm{m^3/s^2}$
(3)	$t_k = t - t_{0e}$	Time from ephemeris epoch
(4)	$M_k = M_0 + n(t_k)$	Mean anomaly
(5)	$M_k = E_k - e\sin E_k$	Eccentric anomaly (must be solved iteratively for E_k)
(6)	$\sin\nu_k = \frac{\sqrt{1-e^2}\sin E_k}{1 - e\cos E_k}$ $\cos\nu_k = \frac{\cos E_k - e}{1 - e\cos E_k}$	True anomaly
(7)	$\phi_k = \nu_k + \omega$	Argument of latitude
(8)	$\delta\phi_k = C_{us}\sin(2\phi_k) + C_{uc}\cos(2\phi_k)$	Argument of latitude correction
(9)	$\delta r_k = C_{rs}\sin(2\phi_k) + C_{rc}\cos(2\phi_k)$	Radius correction
(10)	$\delta i_k = C_{is}\sin(2\phi_k) + C_{ic}\cos(2\phi_k)$	Inclination correction
(11)	$u_k = \phi_k + \delta\phi_k$	Corrected argument of latitude
(12)	$r_k = a(1 - e\cos E_k) + \delta r_k$	Corrected radius
(13)	$i_k = i_0 + (di/dt)t_k + \delta i_k$	Corrected inclination
(14)	$\Omega_k = \Omega_0 + \left(\dot{\Omega} - \dot{\Omega}_e\right)(t_k) - \dot{\Omega}_e t_{0e}$	Corrected longitude of node
(15)	$x_p = r_k \cos u_k$	In-plane x position
(16)	$y_p = r_k \sin u_k$	In-plane y position
(17)	$x_s = x_p\cos\Omega_\kappa - y_p\cos i_k\sin\Omega_\kappa$	ECEF x-coordinate
(18)	$y_s = x_p\sin\Omega_k + y_p\cos i_k\cos\Omega_k$	ECEF y-coordinate
(19)	$z_s = y_p\sin i_k$	ECEF z-coordinate

2.4 Position Determination Using PRN Codes

GPS satellite transmissions utilize direct sequence spread spectrum (DSSS) modulation. DSSS provides the structure for the transmission of ranging signals and essential navigation data, such as satellite ephemerides and satellite health. The ranging signals are PRN codes that binary phase shift key (BPSK) modulate the satellite carrier frequencies. These codes look like and have spectral properties similar to random binary sequences but are actually deterministic. A simple example of a short PRN code sequence is shown in Figure 2.14. These codes have a predictable pattern, which is periodic and can be replicated by a suitably equipped receiver. At the time of this writing, each GPS satellite broadcasted two types of PRN ranging codes: a "short" coarse/acquisition (C/A)-code and a "long" precision (P)-code. (Additional signals are planned to be broadcast. They are described in Chapter 4.) The C/A code has a 1-ms period and repeats constantly, whereas the P-code satellite transmission is a 7-day sequence that repeats approximately every Saturday/Sunday midnight. Presently, the P-code is encrypted. This encrypted code is denoted as the Y-code. The Y-code is accessible only to PPS users through cryptography. Further details regard-

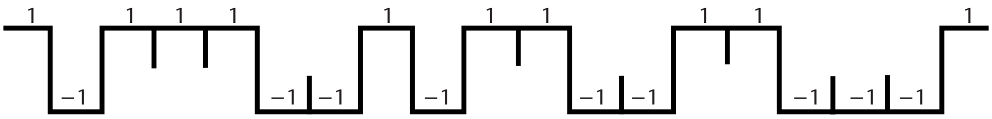


Figure 2.14 PRN ranging code.

ing PRN code properties, frequency generation, and associated modulation processes are contained in Chapter 4.

2.4.1 Determining Satellite-to-User Range

Earlier, we examined the theoretical aspects of using satellite ranging signals and multiple spheres to solve for user position in three dimensions. That example was predicated on the assumption that the receiver clock was perfectly synchronized to system time. In actuality, this is generally not the case. Prior to solving for three-dimensional user position, we will examine the fundamental concepts involving satellite-to-user range determination with nonsynchronized clocks and PRN codes. There are a number of error sources that affect range measurement accuracy (e.g., measurement noise and propagation delays); however, these can generally be considered negligible when compared to the errors experienced from nonsynchronized clocks. Therefore, in our development of basic concepts, errors other than clock offset are omitted. Extensive treatment of these error sources is provided in Section 7.2.

In Figure 2.15, we wish to determine vector **u**, which represents a user receiver's position with respect to the ECEF coordinate system origin. The user's position coordinates x_u, y_u, z_u are considered unknown. Vector **r** represents the vector offset from the user to the satellite. The satellite is located at coordinates x_s, y_s, z_s within the

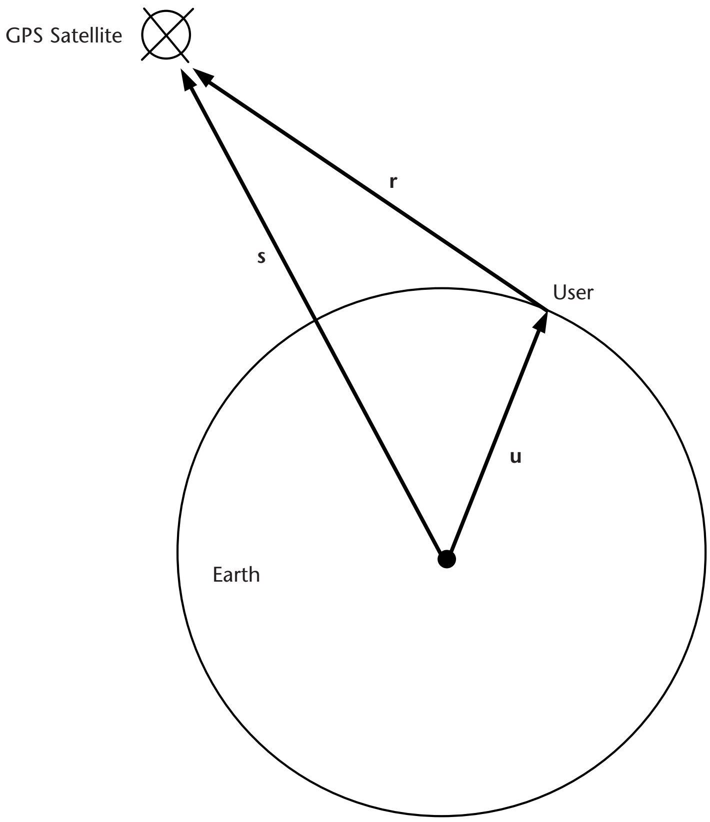


Figure 2.15 User position vector representation.

ECEF Cartesian coordinate system. Vector **s** represents the position of the satellite relative to the coordinate origin. Vector **s** is computed using ephemeris data broadcast by the satellite. The satellite-to-user vector **r** is

$$\mathbf{r} = \mathbf{s} - \mathbf{u} \tag{2.15}$$

The magnitude of vector **r** is

$$\|\mathbf{r}\| = \|\mathbf{s} - \mathbf{u}\| \tag{2.16}$$

Let r represent the magnitude of **r**

$$r = \|\mathbf{s} - \mathbf{u}\| \tag{2.17}$$

The distance r is computed by measuring the propagation time required for a satellite-generated ranging code to transit from the satellite to the user receiver antenna. The propagation time measurement process is illustrated in Figure 2.16. As an example, a specific code phase generated by the satellite at t_1 arrives at the

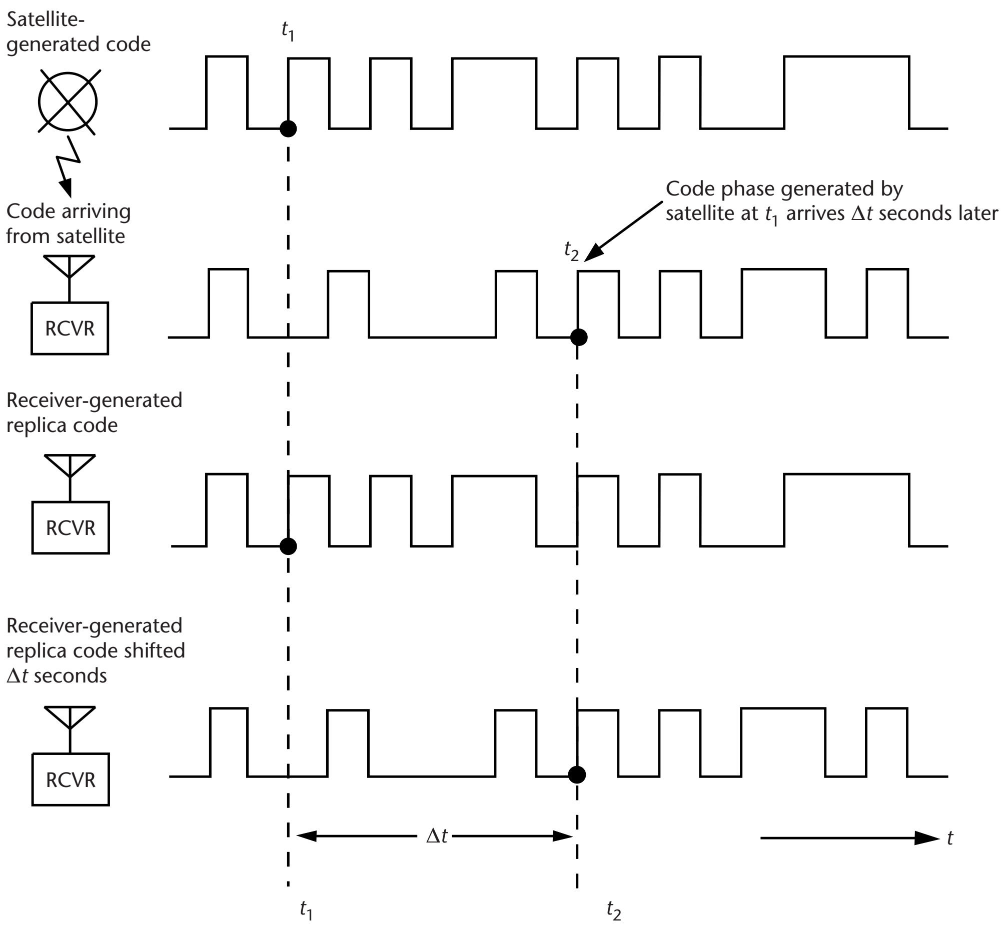


Figure 2.16 Use of replica code to determine satellite code transmission time.

receiver at t_2. The propagation time is represented by Δt. Within the receiver, an identical coded ranging signal is generated at t, with respect to the receiver clock. This replica code is shifted in time until it achieves correlation with the received satellite-generated ranging code. If the satellite clock and the receiver clock were perfectly synchronized, the correlation process would yield the true propagation time. By multiplying this propagation time, Δt, by the speed of light, the true (i.e., geometric) satellite-to-user distance can be computed. We would then have the ideal case described in Section 2.1.2.1. However, the satellite and receiver clocks are generally not synchronized.

The receiver clock will generally have a bias error from system time. Further, satellite frequency generation and timing is based on a highly accurate free running cesium or rubidium atomic clock, which is typically offset from system time. Thus, the range determined by the correlation process is denoted as the pseudorange ρ. The measurement is called *pseudorange* because it is the range determined by multiplying the signal propagation velocity, c, by the time difference between two nonsynchronized clocks (the satellite clock and the receiver clock). The measurement contains (1) the geometric satellite-to-user range, (2) an offset attributed to the difference between system time and the user clock, and (3) an offset between system time and the satellite clock. The timing relationships are shown in Figure 2.17, where:

T_s = System time at which the signal left the satellite

T_u = System time at which the signal reached the user receiver

δt = Offset of the satellite clock from system time [advance is positive; retardation (delay) is negative]

t_u = Offset of the receiver clock from system time

$T_s + \delta t$ = Satellite clock reading at the time that the signal left the satellite

$T_u + t_u$ = User receiver clock reading at the time the signal reached the user receiver

c = speed of light

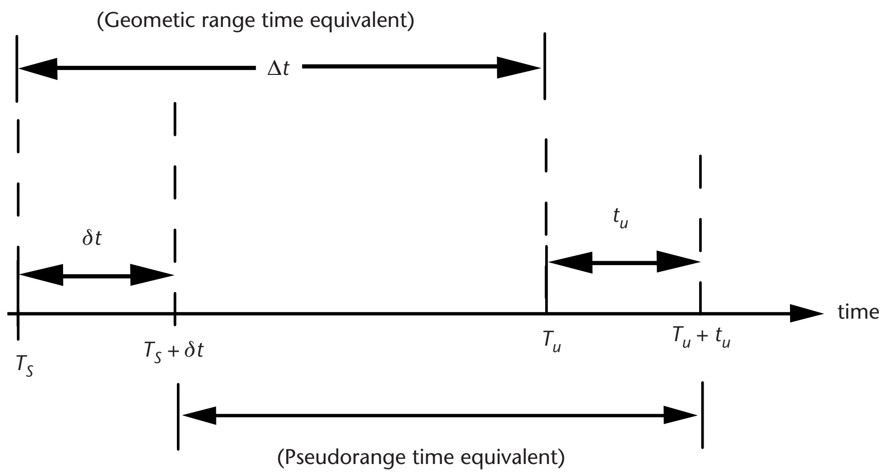


Figure 2.17 Range measurement timing relationships.

$$\textit{Geometric range, } r = c(T_u - T_s) = c\Delta t$$

$$\begin{aligned}\textit{Pseudorange, } \rho &= c\left[(T_u + t_u) - (T_S + \delta t)\right] \\ &= c(T_u - T_S) + c(t_u - \delta t) \\ &= r + c(t_u - \delta t)\end{aligned}$$

Therefore, (2.15) can be rewritten as:

$$\rho - c(t_u - \delta t) = \|\mathbf{s} - \mathbf{u}\|$$

where t_u represents the advance of the receiver clock with respect to system time, δt represents the advance of the satellite clock with respect to system time, and c is the speed of light.

The satellite clock offset from system time, δt, is composed of bias and drift contributions. The GPS ground-monitoring network determines corrections for these offset contributions and transmits the corrections to the satellites for rebroadcast to the users in the navigation message. These corrections are applied within the user receiver to synchronize the transmission of each ranging signal to system time. Therefore, we assume that this offset is compensated for and no longer consider δt an unknown. (There is some residual offset, which is treated in Section 7.2.1, but in the context of this discussion we assume that this is negligible.) Hence, the preceding equation can be expressed as

$$\rho - ct_u = \|\mathbf{s} - \mathbf{u}\| \tag{2.18}$$

2.4.2 Calculation of User Position

In order to determine user position in three dimensions (x_u, y_u, z_u) and the offset t_u, pseudorange measurements are made to four satellites resulting in the system of equations

$$\rho_j = \|\mathbf{s}_j - \mathbf{u}\| + ct_u \tag{2.19}$$

where j ranges from 1 to 4 and references the satellites. Equation (2.19) can be expanded into the following set of equations in the unknowns x_u, y_u, z_u, and t_u:

$$\rho_1 = \sqrt{(x_1 - x_u)^2 + (y_1 - y_u)^2 + (z_1 - z_u)^2} + ct_u \tag{2.20}$$

$$\rho_2 = \sqrt{(x_2 - x_u)^2 + (y_2 - y_u)^2 + (z_2 - z_u)^2} + ct_u \tag{2.21}$$

$$\rho_3 = \sqrt{(x_3 - x_u)^2 + (y_3 - y_u)^2 + (z_3 - z_u)^2} + ct_u \tag{2.22}$$

$$\rho_4 = \sqrt{(x_4 - x_u)^2 + (y_4 - y_u)^2 + (z_4 - z_u)^2} + ct_u \tag{2.23}$$

where x_j, y_j, and z_j denote the jth satellite's position in three dimensions.

These nonlinear equations can be solved for the unknowns by employing either (1) closed-form solutions [19–22], (2) iterative techniques based on linearization, or (3) Kalman filtering. (Kalman filtering provides a means for improving PVT estimates based on optimal processing of time sequence measurements and is described in Sections 7.3.5 and 9.1.3.) Linearization is illustrated in the following paragraphs. (The following development regarding linearization is based on a similar development in [23].) If we know approximately where the receiver is, then we can denote the offset of the true position (x_u, y_u, z_u) from the approximate position ($\hat{x}_u, \hat{y}_u, \hat{z}_u$) by a displacement ($\Delta x_u, \Delta y_u, \Delta z_u$). By expanding (2.20) to (2.23) in a Taylor series about the approximate position, we can obtain the position offset (Δx_u, Δy_u, Δz_u) as linear functions of the known coordinates and pseudorange measurements. This process is described next.

Let a single pseudorange be represented by

$$\begin{aligned}\rho_j &= \sqrt{\left(x_j - x_u\right)^2 + \left(y_j - y_u\right)^2 + \left(z_j - z_u\right)^2} + ct_u \\ &= f(x_u, y_u, z_u, t_u)\end{aligned} \tag{2.24}$$

Using the approximate position location ($\hat{x}_u, \hat{y}_u, \hat{z}_u$) and time bias estimate $\hat{t}_u$, an approximate pseudorange can be calculated:

$$\begin{aligned}\hat{\rho}_j &= \sqrt{\left(x_j - \hat{x}_u\right)^2 + \left(y_j - \hat{y}_u\right)^2 + \left(z_j - \hat{z}_u\right)^2} + c\hat{t}_u \\ &= f\left(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u\right)\end{aligned} \tag{2.25}$$

As stated earlier, the unknown user position and receiver clock offset is considered to consist of an approximate component and an incremental component:

$$\begin{aligned}x_u &= \hat{x}_u + \Delta x_u \\ y_u &= \hat{y}_u + \Delta y_u \\ z_u &= \hat{z}_u + \Delta z_u \\ t_u &= \hat{t}_u + \Delta t_u\end{aligned} \tag{2.26}$$

Therefore, we can write

$$f(x_u, y_u, z_u, t_u) = f\left(\hat{x}_u + \Delta x_u, \hat{y}_u + \Delta y_u, \hat{z}_u + \Delta z_u, \hat{t}_u + \Delta t_u\right)$$

This latter function can be expanded about the approximate point and associated predicted receiver clock offset ($\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u$) using a Taylor series:

$$\begin{aligned}&f\left(\hat{x}_u + \Delta x_u, \hat{y}_u + \Delta y_u, \hat{z}_u + \Delta z_u, \hat{t}_u + \Delta t_u\right) = f\left(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u\right) \\ &+ \frac{\partial f\left(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u\right)}{\partial \hat{x}_u} \Delta x_u + \frac{\partial f\left(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u\right)}{\partial \hat{y}_u} \Delta y_u \\ &+ \frac{\partial f\left(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u\right)}{\partial \hat{z}_u} \Delta z_u + \frac{\partial f\left(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u\right)}{\partial \hat{t}_u} \Delta t_u + \ldots\end{aligned} \tag{2.27}$$

The expansion has been truncated after the first-order partial derivatives to eliminate nonlinear terms. The partials derivatives evaluate as follows:

$$\begin{aligned}
\frac{\partial f\left(\hat{x}_u,\hat{y}_u,\hat{z}_u,\hat{t}_u\right)}{\partial \hat{x}_u} &= -\frac{x_j - \hat{x}_u}{\hat{r}_j} \\
\frac{\partial f\left(\hat{x}_u,\hat{y}_u,\hat{z}_u,\hat{t}_u\right)}{\partial \hat{y}_u} &= -\frac{y_j - \hat{y}_u}{\hat{r}_j} \\
\frac{\partial f\left(\hat{x}_u,\hat{y}_u,\hat{z}_u,\hat{t}_u\right)}{\partial \hat{z}_u} &= -\frac{z_j - \hat{z}_u}{\hat{r}_j} \\
\frac{\partial f\left(\hat{x}_u,\hat{y}_u,\hat{z}_u,\hat{t}_u\right)}{\partial \hat{t}_u} &= c
\end{aligned} \tag{2.28}$$

where

$$\hat{r}_j = \sqrt{\left(x_j - \hat{x}_u\right)^2 + \left(y_j - \hat{y}_u\right)^2 + \left(z_j - \hat{z}_u\right)^2}$$

Substituting (2.25) and (2.28) into (2.27) yields

$$\rho_j = \hat{\rho}_j - \frac{x_j - \hat{x}_u}{\hat{r}_j}\Delta x_u - \frac{y_j - \hat{y}_u}{\hat{r}_j}\Delta y_u - \frac{z_j - \hat{z}_u}{\hat{r}_j}\Delta z_u + ct_u \tag{2.29}$$

We have now completed the linearization of (2.24) with respect to the unknowns Δx_u, Δy_u, Δz_u, and Δt_u. (It is important to remember that we are neglecting secondary error sources such as Earth rotation compensation, measurement noise, propagation delays, and relativistic effects, which are treated in detail in Section 7.2.)

Rearranging this expression with the known quantities on the left and unknowns on right yields

$$\hat{\rho}_j - \rho_j = \frac{x_j - \hat{x}_u}{\hat{r}_j}\Delta x_u + \frac{y_j - \hat{y}_u}{\hat{r}_j}\Delta y_u - \frac{z_j - \hat{z}_u}{\hat{r}_j}\Delta z_u - ct_u \tag{2.30}$$

For convenience, we will simplify the previous equation by introducing new variables where

$$\begin{aligned}
\Delta\rho &= \hat{\rho}_j - \rho_j \\
a_{xj} &= \frac{x_j - \hat{x}_u}{\hat{r}_j} \\
a_{yj} &= \frac{y_j - \hat{y}_u}{\hat{r}_j} \\
a_{zy} &= \frac{z_j - \hat{z}_u}{\hat{r}_j}
\end{aligned} \tag{2.31}$$

The a_{xj}, a_{yj}, and a_{zj} terms in (2.31) denote the direction cosines of the unit vector pointing from the approximate user position to the jth satellite. For the jth satellite, this unit vector is defined as

$$\mathbf{a}_j = \left(a_{xj}, a_{yj}, a_{zj}\right)$$

Equation (2.30) can be rewritten more simply as

$$\Delta\rho_j = a_{xj}\Delta x_u + a_{yj}\Delta y_u + a_{zj}\Delta z_u - c\Delta t_u$$

We now have four unknowns: Δx_u, Δy_u, Δz_u, and Δt_u, which can be solved for by making ranging measurements to four satellites. The unknown quantities can be determined by solving the set of linear equations that follow:

$$\begin{aligned}
\Delta\rho_1 &= a_{x1}\Delta x_u + a_{y1}\Delta y_u + a_{z1}\Delta z_u - c\Delta t_u \\
\Delta\rho_2 &= a_{x2}\Delta x_u + a_{y2}\Delta y_u + a_{z2}\Delta z_u - c\Delta t_u \\
\Delta\rho_3 &= a_{x3}\Delta x_u + a_{y3}\Delta y_u + a_{z3}\Delta z_u - c\Delta t_u \\
\Delta\rho_4 &= a_{x4}\Delta x_u + a_{y4}\Delta y_u + a_{z4}\Delta z_u - c\Delta t_u
\end{aligned} \tag{2.32}$$

These equations can be put in matrix form by making the definitions

$$\Delta\boldsymbol{\rho} = \begin{bmatrix} \Delta\rho_1 \\ \Delta\rho_2 \\ \Delta\rho_3 \\ \Delta\rho_4 \end{bmatrix} \quad \mathbf{H} = \begin{bmatrix} a_{x1} & a_{y1} & a_{z1} & 1 \\ a_{x2} & a_{y2} & a_{z2} & 1 \\ a_{x3} & a_{y3} & a_{z3} & 1 \\ a_{x4} & a_{y4} & a_{z4} & 1 \end{bmatrix} \quad \Delta\mathbf{x} = \begin{bmatrix} \Delta x_u \\ \Delta y_u \\ \Delta z_u \\ -c\Delta t_u \end{bmatrix}$$

One obtains, finally,

$$\Delta\boldsymbol{\rho} = \mathbf{H}\Delta\mathbf{x} \tag{2.33}$$

which has the solution

$$\Delta\mathbf{x} = \mathbf{H}^{-1}\Delta\boldsymbol{\rho} \tag{2.34}$$

Once the unknowns are computed, the user's coordinates x_u, y_u, z_u and the receiver clock offset t_u are then calculated using (2.26). This linearization scheme will work well as long as the displacement (Δx_u, Δy_u, Δz_u) is within close proximity of the linearization point. The acceptable displacement is dictated by the user's accuracy requirements. If the displacement does exceed the acceptable value, this process is reiterated with $\hat{\rho}$ being replaced by a new estimate of pseudorange based on the calculated point coordinates x_u, y_u, and z_u. In actuality, the true user-to-satellite measurements are corrupted by uncommon (i.e., independent) errors, such as measurement noise, deviation of the satellite path from the reported ephemeris, and multipath. These errors translate to errors in the components of vector $\Delta\mathbf{x}$, as shown here:

$$\boldsymbol{\epsilon}_{\mathbf{x}} = \mathbf{H}^{-1}\boldsymbol{\epsilon}_{\mathbf{meas}} \tag{2.35}$$

where $\boldsymbol{\epsilon}_{meas}$ is the vector containing the pseudorange measurement errors and $\boldsymbol{\epsilon}_x$ is the vector representing errors in the user position and receiver clock offset.

The error contribution $\boldsymbol{\epsilon}_x$ can be minimized by making measurements to more than four satellites, which will result in an overdetermined solution set of equations similar to (2.33). Each of these redundant measurements will generally contain independent error contributions. Redundant measurements can be processed by least squares estimation techniques that obtain improved estimates of the unknowns. Various versions of this technique exist and are usually employed in today's receivers, which generally employ more than four user-to-satellite measurements to compute user PVT. Appendix A provides an introduction to least squares techniques.

CHAPTER 3

GPS System Segments

Arthur J. Dorsey and Willard A. Marquis
Lockheed Martin Corporation

Peter M. Fyfe
The Boeing Company

Elliott D. Kaplan and Lawrence F. Wiederholt
The MITRE Corporation

3.1 Overview of the GPS System

GPS is comprised of three segments: satellite constellation, ground-control/monitoring network, and user receiving equipment. Formal GPS JPO programmatic terms for these components are space, control, and user equipment segments, respectively. The satellite constellation is the set of satellites in orbit that provide the ranging signals and data messages to the user equipment. The control segment (CS) tracks and maintains the satellites in space. The CS monitors satellite health and signal integrity and maintains the orbital configuration of the satellites. Furthermore, the CS updates the satellite clock corrections and ephemerides as well as numerous other parameters essential to determining user PVT. Finally, the user receiver equipment (i.e., user segment) performs the navigation, timing, or other related functions (e.g., surveying). An overview of each system segment is provided next, followed by further elaboration on each segment starting in Section 3.2.

3.1.1 Space Segment Overview

The space segment is the constellation of satellites from which users make ranging measurements. The SVs (i.e., satellites) transmit a PRN-coded signal from which the ranging measurements are made. This concept makes GPS a passive system for the user with signals only being transmitted and the user passively receiving the signals. Thus, an unlimited number of users can simultaneously use GPS. A satellite's transmitted ranging signal is modulated with data that includes information that defines the position of the satellite. An SV includes payloads and vehicle control subsystems. The primary payload is the navigation payload used to support the GPS PVT mission; the secondary payload is the nuclear detonation (NUDET) detection system, which supports detection and reporting of Earth-based radiation phenomena.

The vehicle control subsystems perform such functions as maintaining the satellite pointing to Earth and the solar panels pointing to the Sun.

3.1.2 Control Segment (CS) Overview

The CS is responsible for maintaining the satellites and their proper functioning. This includes maintaining the satellites in their proper orbital positions (called stationkeeping) and monitoring satellite subsystem health and status. The CS also monitors the satellite solar arrays, battery power levels, and propellant levels used for maneuvers. Furthermore, the CS activates spare satellites (if available) to maintain system availability. The CS updates each satellite's clock, ephemeris, and almanac and other indicators in the navigation message at least once per day. Updates are more frequently scheduled when improved navigation accuracies are required. (Frequent clock and ephemeris updates result in reducing the space and control contributions to range measurement error. Further elaboration on the effects of frequent clock and ephemeris updates is provided in Sections 3.3.1.4 and 7.2).

The ephemeris parameters are a precise fit to the GPS satellite orbits and are valid only for a time interval of 4 hours with the once-per-day normal upload schedule. Depending on the satellite block, the navigation message data can be stored for a minimum of 14 days to a maximum of a 210-day duration in intervals of 4 hours or 6 hours for uploads as infrequent as once per two weeks and intervals of greater than 6 hours in the event that an upload cannot be provided for over 2 weeks. The almanac is a reduced precision subset of the ephemeris parameters. The almanac consists of 7 of the 15 ephemeris orbital parameters. Almanac data is used to predict the approximate satellite position and aid in satellite signal acquisition. Furthermore, the CS resolves satellite anomalies, controls SA and AS (see Sections 1.3.1 and 7.2.1), and collects pseudorange and carrier phase measurements at the remote monitor stations to determine satellite clock corrections, almanac, and ephemeris. To accomplish these functions, the CS is comprised of three different physical components: the master control station (MCS), monitor stations, and the ground antennas, each of which is described in more detail in Section 3.3.

3.1.3 User Segment Overview

The user receiving equipment comprises the user segment. Each set of equipment is typically referred to as a *GPS receiver*, which processes the L-band signals transmitted from the satellites to determine user PVT. While PVT determination is the most common use, receivers are designed for other applications, such as computing user platform attitude (i.e., heading, pitch, and roll) or as a timing source. Section 3.4 provides further discussion on the user segment.

3.2 Space Segment Description

The space segment has two principal aspects: One aspect is the constellation of satellites in terms of the orbits and positioning within the orbits. The other aspect is the features of the satellites that occupy each orbital slot. Each aspect is described next.

3.2.1 GPS Satellite Constellation Description

The U.S. government baseline configuration for the constellation consists of 24 satellites. Within this configuration, the satellites are positioned in six Earth-centered orbital planes with four satellites in each plane. The nominal orbital period of a GPS satellite is one-half of a sidereal day or 11 hours, 58 minutes [1]. The orbits are nearly circular and equally spaced around the equator at a 60° separation with a nominal inclination relative to the equatorial plane of 55°. Figure 3.1 depicts the GPS constellation. The orbital radius (i.e., nominal distance from the center of mass of the Earth to the satellite) is approximately 26,600 km. This satellite constellation provides a 24-hour global user navigation and time determination capability. Figure 3.2 presents the satellite orbits in a planar projection referenced to the epoch time of 0000h July 1, 1993 UTC (USNO). Thinking of an orbit as a ring, this figure opens each orbit and lays it flat on a plane. Similarly, for the Earth's equator, it is like a ring that has been opened and laid on a flat surface. The slope of each orbit represents its inclination with respect to the Earth's equatorial plane, which is nominally 55°.

The orbital plane locations with respect to the Earth are defined by the longitude of the ascending node, while the location of the satellite within the orbital plane is defined by the mean anomaly. The longitude of the ascending node is the point of intersection of each orbital plane with the equatorial plane. The Greenwich meridian is the reference point where the longitude of the ascending node has the value of zero. Mean anomaly is the angular position of each satellite within the orbit, with the Earth's equator being the reference or point with a zero value of mean anomaly. It can be observed that the relative phasing between most satellites in adjoining orbits is approximately 40°. The Keplerian parameters for the 24-SV constellation are defined in Section 2.3.1.

The orbital slot assignments of this baseline design are contained in [2] and are provided in Table 3.1. (Note that RAAN is the Right Ascension of the Ascending Node, as defined in Section 2.3.1.)

The remaining reference orbit values (with tolerances) are:

Figure 3.1 GPS satellite constellation. (*Source:* Lockheed Martin Corp. Reprinted with permission.)

3.3 Control Segment

The control segment (CS) is responsible for monitoring, commanding, and controlling the GPS satellite constellation. Functionally, the CS monitors the downlink L-band navigation signals, updates the navigation messages, and resolves satellite anomalies. Additionally, the CS monitors each satellite's state of health, manages tasks associated with satellite stationkeeping maneuvers and battery recharging, and commands the satellite payloads, as required [11].

The major elements of the CS consist of the MCS, L-band monitor stations, and S-band ground antennas. The primary CS functions are performed at the MCS, under the operation of the U.S. Air Force Space Command, Second Space Operation Squadron (2SOPS), located at Schriever Air Force Base (AFB) in Colorado Springs, Colorado. It provides continuous GPS services, 24 hours per day, 7 days a week, and serves as the mission control center for GPS operations. A backup MCS, located at a contractor facility in Gaithersburg, Maryland, provides redundancy of the MCS. The major elements of the CS and their functional allocation are shown in Figure 3.14.

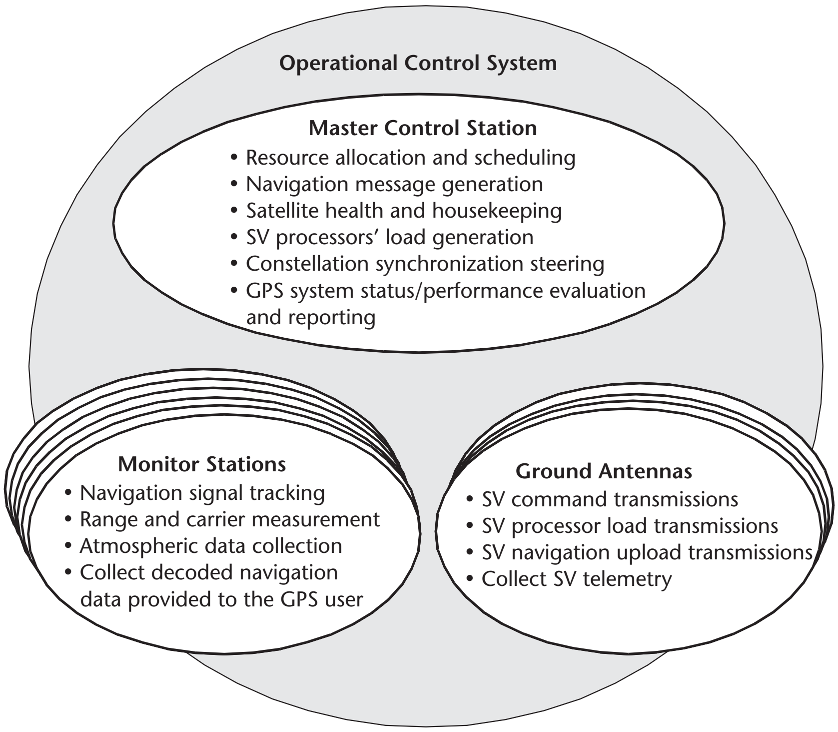


Figure 3.14 CS overview.

3.4 User Segment

The user receiving equipment, typically referred to as a GPS receiver, processes the L-band signals transmitted from the satellites to determine PVT. Technology trends in component miniaturization and large-scale manufacturing have led to a proliferation of low-cost GPS receiver components. GPS receivers are embedded in many of the items we use in our daily lives. These items include cellular telephones, PDAs, and automobiles. This is in contrast to the initial receiving sets manufactured in the mid-1970s as part of the system concept validation phase. These first receivers were primarily analog devices for military applications and were large, bulky, and heavy. Today, receivers take on many form factors, including chipsets, handheld units, and Industry Standard Architecture (ISA) compatible cards. In fact, there are many single-chip GPS receivers that have leveraged low-voltage bipolar complementary metal oxide semiconductor (BiCMOS) processes and power-management techniques to meet the need for small size and low battery drain of handheld devices. Selection of a GPS receiver depends on the user's application (e.g., civilian versus military, platform dynamics, and shock and vibration environment). Following a description of a typical receiver's components, selection criteria are addressed. Detailed information regarding GPS receiver architectures and integrations for cellular telephone and automotive applications is contained in Chapter 9.

3.4.1 GPS Set Characteristics

A block diagram of a GPS receiving set is shown in Figure 3.23. The GPS set consists of five principal components: antenna, receiver, processor, input/output (I/O) device such as a control display unit (CDU), and a power supply.

New civil signals L2C and L5 have null-to-null bandwidths of 2.046 MHz and 20.46 MHz, respectively. The military M code can be processed within the existing L1 and L2 24-MHz bandwidths. Since M code signal power is defined within a 30.69-MHz band around the center frequency, approximately 92% of this power is within the 24-MHz band. (GPS signal characteristics are contained in Chapter 4.)

The addition of new signals (M code, L1C, L2C, and L5) will require new antennas for some users. For example, those utilizing L1 C/A code and L2C will need a dual-band antenna. (Dual frequency measurements enable determination of the ionospheric delay and provide robustness to interference. Ionospheric delay determination and compensation are discussed in Chapter 7.) SOL signal users that require operation in the ARNS bands will need antennas to receive C/A code on L1 and the L5 signal on L5. At the time of this writing, RTCA was developing aviation standards for a dual-band L1/L5 antenna. Some receivers may be tri-band. That is, they will receive and process the signals broadcast on all three GPS frequencies, L1, L2, and L5, which will require a tri-band antenna. Reference [33] provides details on one approach for a tri-band (L1/L2 M code and L5) antenna design.

Antenna designs vary from helical coils to thin microstrip (i.e., patch) antennas. High-dynamic aircraft prefer low-profile, low–air resistance patch antennas, whereas land vehicles can tolerate a larger antenna. Antenna selection requires evaluation of such parameters as antenna gain pattern, available mounting area, aerodynamic performance, multipath performance, and stability of the electrical phase center of the antenna [34].

Another issue regarding antenna selection is the need for resistance to interference. (In the context of this discussion, any electronic emission, whether friendly or hostile, that interferes with the reception and processing of GPS signals is considered an interferer.) Some military aircraft employ antenna arrays to form a null in the direction of the interferer. Another technique to mitigate the effects of interference is to employ a beam-steering array. Beam-steering techniques electronically concentrate the antenna gain in the direction of the satellites to maximize link margin. Finally, beam forming combines both nulling and beam steering for interferer mitigation. (References [35–37] provide detailed descriptions of the theory and practical applications of nulling, beam steering, and beam forming.)

3.4.1.2 Receiver

Chapter 5 provides a detailed description of receiver signal acquisition and tracking operation; however, some high-level aspects are described herein to aid our discussion. Two basic receiver types exist today: (1) those that track L1 C/A code and P(Y) code on L1 and L2 and (2) those that only track C/A code. In light of the GPS modernization effort, these are referred to as legacy receivers. Forthcoming military receivers are being referred to as YMCA. That is, they will track L1 C/A, L1 and L2 P(Y), and L1 and L2 M code. The forthcoming civil signals, L1C, L2C, and L5, will require new receivers to be built. It is envisioned that a number of receiver types will be available. Most likely, these will be dual band to achieve ionospheric compensation and increased interference immunity. As mentioned earlier, ARNS band users will require dual band (L1 and L5) receivers and antennas.

Legacy PPS users generally employ sets that track P(Y) code on both L1 and L2. These sets initiate operation with receivers tracking C/A code on L1 and then transition to tracking P(Y) code on both L1 and L2. Y-code tracking occurs only with the aid of cryptographic equipment. (If the satellite signal is encrypted and the receiver does not have the proper cryptographic equipment, the receiver generally defaults to tracking C/A code on L1.) It is anticipated that the forthcoming YMCA receivers will perform a direct acquisition of the M code signal. Following M code acquisition, the receivers will then track M code on both L1 and L2 if the receiver is capable of dual-frequency operation. Otherwise, it will operate on either L1 or L2.

Alternatively, legacy SPS users employ sets that track the C/A code exclusively on L1, since that is the only frequency on which the C/A code is generally broadcast. Forthcoming L1C, L2C, and L5 receivers will track signals on these respective frequencies.

In addition to the receiver types mentioned earlier, there are other variations, such as civilian semicodeless tracking receivers, which track the C/A code on L1 and carrier phase of both the L1 and L2 frequencies. These receivers employ signal-processing techniques that to do not require cryptographic access to the P(Y) code. Utilizing the carrier phase as a measurement observable enables centimeter-level (or even millimeter-level) measurement accuracy. (Carrier-phase measurements are described extensively in Section 8.4.) Most receivers have multiple channels whereby each channel tracks the transmission from a single satellite. A simplified block diagram of a multichannel generic SPS receiver is shown in Figure 3.25. The

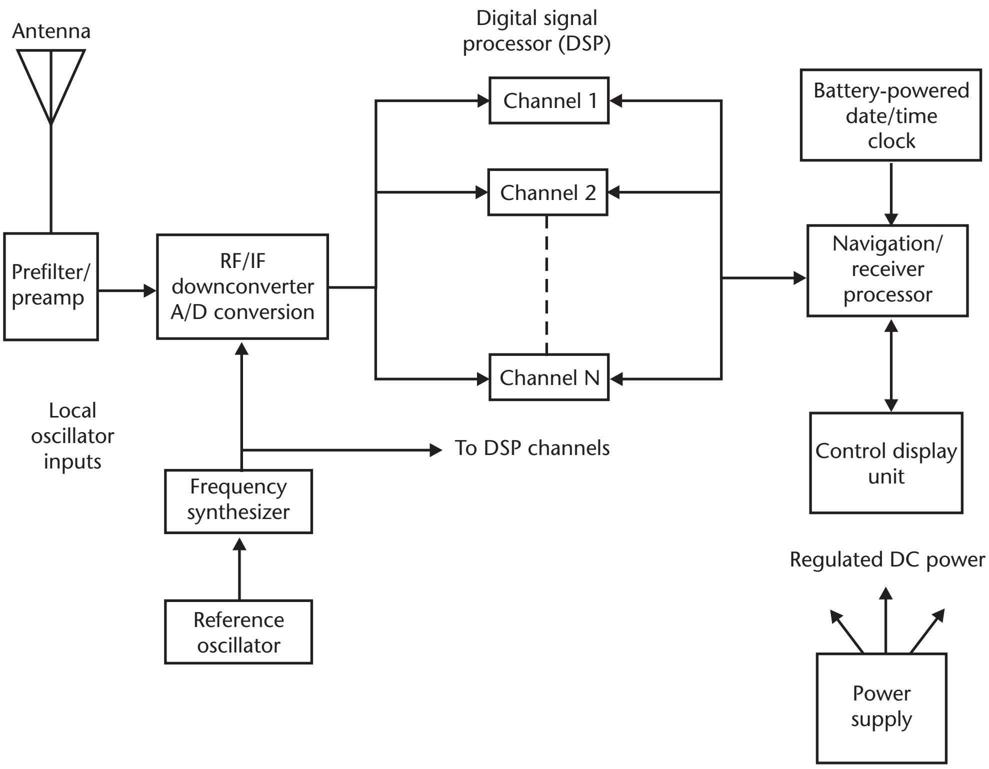


Figure 3.25 Generic SPS receiver.

received RF CDMA satellite signals are usually filtered by a passive bandpass prefilter to reduce out-of-band RF interference.

This is normally followed by a preamplifier. The RF signals are then downconverted to an intermediate frequency (IF). The IF signals are sampled and digitized by an analog to digital (A/D) converter. The A/D sampling rate is typically 2 to 20 times the PRN code chipping rate [1.023 MHz for L1 C/A code and 10.23 MHz for L1 and L2 P(Y) code]. The minimum sampling rate is twice the stopband bandwidth of the codes to satisfy the Nyquist criterion. For L1 C/A code only sets, the stopband bandwidth may be slightly greater than 1 MHz. Alternatively, the stopband bandwidth is slightly more than 10 MHz for P(Y) code sets. Oversampling reduces the receiver sensitivity to A/D quantization noise, thereby reducing the number of bits required in the A/D converter. The samples are forwarded to the digital signal processor (DSP). The DSP contains N parallel channels to simultaneously track the carriers and codes from up to N satellites. (N generally ranges from 8 to 12 in today's receivers.) Each channel contains code and carrier tracking loops to perform code and carrier-phase measurements, as well as navigation message data demodulation. The channel may compute three different satellite-to-user measurement types: pseudoranges, delta ranges (sometimes referred to as delta pseudorange), and integrated Doppler, depending on the implementation. The desired measurements and demodulated navigation message data are forwarded to the processor.

Note that GPS receivers designed for use in handheld devices need to be power efficient. Depending on the implementation, these receivers may trade off susceptibility to high-power in-band interferers to achieve minimum power supply (e.g., battery) drain. High dynamic range receiver front ends are needed in interference-resistant receivers, and the necessary components (e.g., amplifiers and mixers with high intermodulation product levels) require high bias voltage levels.

3.4.1.3 Navigation/Receiver Processor

A processor is generally required to control and command the receiver through its operational sequence, starting with channel signal acquisition and followed by signal tracking and data collection. (Some GPS sets have an integral processing capability within the channel circuitry to perform these signal-processing functions.) In addition, the processor may also form the PVT solution from the receiver measurements. In some applications, a separate processor may be dedicated to the computation of both PVT and associated navigation functions. Most processors provide an independent PVT solution on a 1-Hz basis. However, receivers designated for autoland aircraft precision approach and other high-dynamic applications normally require computation of independent PVT solutions at a minimum of 5 Hz. The formulated PVT solution and other navigation-related data is forwarded to the I/O device.

3.4.1.4 I/O Device

The I/O device is the interface between the GPS set and the user. I/O devices are of two basic types: integral or external. For many applications, the I/O device is a

CDU. The CDU permits operator data entry, displays status and navigation solution parameters, and usually accesses numerous navigation functions, such as waypoint entry and time to go. Most handheld units have an integral CDU. Other installations, such as those onboard an aircraft or ship, may have the I/O device integrated with existing instruments or control panels. In addition to the user and operator interface, applications such as integration with other sensors (e.g., INS) require a digital data interface to input and output data. Common interfaces are ARINC 429, MIL-STD-1553B, RS-232, and RS-422.

3.4.1.5 Power Supply

The power supply can be either integral, external, or a combination of the two. Typically, alkaline or lithium batteries are used for integral or self-contained implementations, such as handheld portable units; whereas an existing power supply is normally used in integrated applications, such as a board-mounted receiver installed within a server to provide accurate time. Airborne, automotive, and shipboard GPS set installations normally use platform power but typically have built-in power converters (ac to dc or dc to dc) and regulators. There usually is an internal battery to maintain data stored in volatile random access memory (RAM) integrated circuits (ICs) and to operate a built-in timepiece (date/time clock) in the event platform power is disconnected.

3.4.2 GPS Receiver Selection

At the time of this writing, there were well over 100 GPS set manufacturers in the United States and abroad. While some, like SiRF, offer a few different chip set receivers for integration with other electronic functions, other companies like GARMIN and Trimble Navigation have many different end products ranging from handhelds to automobile and aircraft navigators to complex survey receivers. GPS receiver selection is dependent on user application. The intended application strongly influences receiver design, construction, and capability. For each application, numerous environmental, operational, and performance parameters must be examined. A sampling of these parameters follows:

- What are the shock and vibration requirements, temperature and humidity extremes, as well as atmospheric salt content?
- If the receiver is to be used by government or military personnel, PPS operation may be required. PPS operation usually dictates that a dual-frequency set with a cryptographic capability is needed.
- The necessary independent PVT update rate must be determined. As an example, this rate is different for aircraft precision approach than it is for marine oil tanker guidance.
- Will the receiver have to operate in a high-multipath environment (i.e., near buildings or on an aircraft where satellite signals are reflected by various fuselage surfaces)? If so, multipath mitigation signal-processing techniques may be required. (Detailed descriptions of multipath and multipath-mitigation

techniques are contained in Chapter 6. The contribution to the GPS error budget is described in Chapter 7.)

- Under what type of dynamic conditions (e.g., acceleration and velocity) will the set have to operate? GPS sets for fighter aircraft applications are designed to maintain full performance even while experiencing multiple “Gs” of acceleration, whereas sets designated for surveying are not normally designed for severe dynamic environments.
- Is a DGPS capability required? (DGPS is an accuracy-enhancement technique covered in Chapter 8.) DGPS provides greater accuracy than stand-alone PPS and SPS. Most receivers are manufactured with a DGPS capability.
- Does the application require reception of the geostationary satellite-based overlay service referred to as SBAS broadcasting satellite integrity, ranging, and DGPS information? (SBAS is discussed in Chapter 8.)
- Waypoint storage capability as well as the number of routes and legs need to be assessed.
- Does the GPS set have to operate in an environment that requires enhanced interference rejection capabilities? Chapter 6 describes several techniques to achieve this.
- If the receiver has to be interfaced with an external system, does the proper I/O hardware and software exist? An example would be a user who requires a blended solution consisting of GPS and other sensors, such as an IMU and vision system.
- In terms of data input and display features, does the receiver require an external or integral CDU capability? Some aircraft and ships use *repeater* units such that data can be entered or extracted from various physical locations. Display requirements such as sunlight-readable or night-vision-goggle-compatible must be considered.
- Are local datum conversions required, or is WGS-84 sufficient? If so, does the receiver contain the proper transformations?
- Is portability for field use required?
- Economics, physical size, and power consumption must also be considered.

As stated earlier, these are only a sampling of GPS set selection parameters. One must carefully review the requirements of the user application prior to selecting a receiver. In most cases, the selection will be a tradeoff that requires awareness of the impact of any GPS set deficiencies for the intended application.

CHAPTER 4

GPS Satellite Signal Characteristics

Phillip W. Ward
NAVWARD GPS Consulting

John W. Betz and Christopher J. Hegarty
The MITRE Corporation

4.1 Overview

In this chapter, we examine the properties of the GPS satellite signals, including frequency assignment, modulation format, navigation data, and the generation of PRN codes. This discussion is accompanied by a description of received signal power levels as well as their associated autocorrelation characteristics. Cross-correlation characteristics are also described. The chapter is organized as follows. First, background information on modulations that are useful for satellite radionavigation, multiplexing techniques, and general signal characteristics including autocorrelation functions and power spectra are discussed in Section 4.2. Section 4.3 describes the *legacy* GPS signals, defined here as those signals broadcast by the GPS satellites up through the Block IIR SVs. Section 4.4 presents an overview of the GPS navigation data modulated upon the legacy GPS signals. As discussed in Chapter 3, new civil and military signals will be broadcast by the Block IIR-M and later satellites. These new signals are discussed in Section 4.5. Finally, Section 4.6 summarizes the chapter.

4.2 Modulations for Satellite Navigation

4.2.1 Modulation Types

Binary phase shift keying (BPSK) is a simple digital signaling scheme in which an RF carrier is either transmitted "as is" or with a 180º phase shift over successive intervals in time depending on whether a digital 0 or 1 is being conveyed (e.g., see [1]). A BPSK signal, as illustrated in Figure 4.1, can be thought of as the product of two time waveforms—the unmodulated RF carrier and a data waveform that takes on a value of either +1 or –1 for each successive interval of $T_b = 1/R_b$ seconds, where R_b is the data rate in bits per second. The data waveform amplitude for the kth interval of T_b seconds can be generated from the kth data bit to be transmitted using either the mapping $[0, 1] \rightarrow [-1, +1]$ or $[0, 1] \rightarrow [+1, -1]$. In many systems, *forward error cor-*

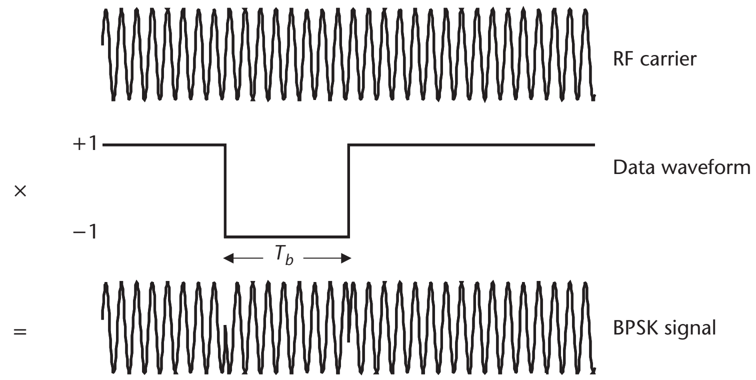


Figure 4.1 BPSK modulation.

rection (FEC) is employed, whereby redundant bits (more than the original information bits) are transmitted over the channel according to some prescribed method, enabling the receiver to detect and correct some errors that may be introduced by noise, interference, or fading. When FEC is employed, common convention is to replace T_b with T_s and R_b with R_s to distinguish data symbols (actually transmitted) from data bits (that contain the information before FEC). The data waveform alone is considered a *baseband* signal, meaning that its frequency content is concentrated around 0 Hz rather than the carrier frequency. Modulation by the RF carrier centers the frequency content of the signal about the carrier frequency, creating what is known as a *bandpass* signal.

Direct sequence spread spectrum (DSSS) is an extension of BPSK or other phase shift keyed modulation used by GPS and other satellite navigation systems discussed in this text. As shown in Figure 4.2, DSSS signaling adds a third component, referred to as a *spreading* or PRN waveform, which is similar to the data waveform but at a much higher symbol rate. This PRN waveform is completely known, at least to the intended receivers. The PRN waveform is often periodic, and the finite sequence of bits used to generate the PRN waveform over one period is referred to as a *PRN*

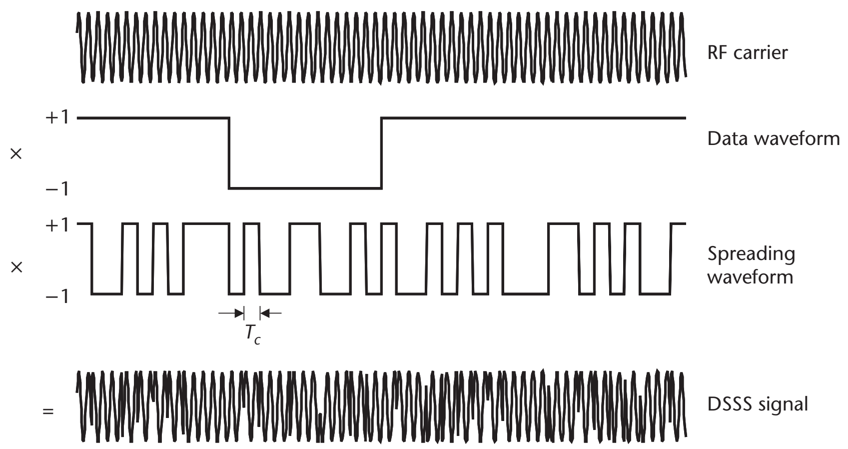


Figure 4.2 DSSS modulation.

sequence or *PRN code*. An excellent overview of PRN codes, including their generation, characteristics, and code families with good properties is provided in [2]. The minimum interval of time between transitions in the PRN waveform is commonly referred to as the *chip period*, T_c; the portion of the PRN waveform over one chip period is known as a *chip* or *spreading symbol*; and the reciprocal of the chip period is known as the *chipping rate*, R_c. The independent time parameter for the PRN waveform is often expressed in units of chips and referred to as *codephase*.

The signal just described is called *spread spectrum,* because of the wider bandwidth occupied by the signal after modulation by the high-rate PRN waveform. In general, the bandwidth is proportional to the chipping rate.

There are three primary reasons why DSSS waveforms are employed for satellite navigation. First and most importantly, the frequent phase inversions in the signal introduced by the PRN waveform enable precise ranging by the receiver. Second, the use of different PRN sequences from a well-designed set enables multiple satellites to transmit signals simultaneously and at the same frequency. A receiver can distinguish among these signals, based on their different codes. For this reason, the transmission of multiple DSSS signals having different spreading sequences on a common carrier frequency is referred to as *code division multiple access* (CDMA). Finally, as detailed in Chapter 6, DSSS provides significant rejection of narrowband interference.

It should be noted that the chip waveform in a DSSS signal does not need to be rectangular (i.e., a constant amplitude over the chip period), as we have assumed earlier. In principle, any shape could be used and different shapes can be used for different chips. Henceforth, we will denote DSSS signals generated using BPSK signaling with rectangular chips as *BPSK-R* signals. Several variations of the basic DSSS signal that employ nonrectangular symbols have been investigated for satellite navigation applications in recent years. *Binary offset carrier* (BOC) signals [3] are generated using DSSS techniques but employ portions of a square wave for the spreading symbols. A generalized treatment of the use of arbitrary binary patterns to generate each spreading symbol is provided in [4]. Spreading symbol shapes, such as raised cosines, whose amplitudes vary over a wide range of values, are used extensively in digital communications. These shapes have also been considered for satellite navigation but to date have not been used for practical reasons. For precise ranging, it is necessary for the satellite and user equipment to be able to faithfully reproduce the spreading waveform, which is facilitated through the use of signals that can be generated using simple digital means. Furthermore, spectral efficiency, which has motivated extensive studies in symbol shaping for communications applications, is generally not a concern for satellite navigation. Finally, DSSS signals with *constant envelope* (e.g., those that employ binary-valued—one magnitude with two possible polarities—spreading symbols) can be efficiently transmitted using switching-class amplifiers.

4.3 Legacy GPS Signals

This section details the legacy GPS navigation signals—that is, those navigation signals that are broadcast by the GPS SVs up through the Block IIR class (see Chapter 3). The legacy GPS SVs transmit navigation signals on two carrier frequencies called L1, the primary frequency, and L2, the secondary frequency. The carrier frequencies are DSSS modulated by spread spectrum codes with unique PRN sequences associated with each SV and by a common navigation data message. All SVs transmit at the same carrier frequencies in a CDMA fashion. In order to track one SV in common view with several other SVs by the CDMA technique, a GPS receiver must replicate the PRN sequence for the desired SV along with the replica carrier signal, including Doppler effects. Two carrier frequencies are required to measure the ionospheric delay, since this delay is related by a scale factor to the difference in signal TOA for the two carrier frequencies. Single frequency users must estimate the ionospheric delay using modeling parameters that are broadcast to the user in the navigation message. (Further information on ionospheric delay compensation is contained in Section 7.2.4.1.) The characteristics of the legacy GPS signals are further explained in the following sections.

4.3.1 Frequencies and Modulation Format

A block diagram that is representative of the SV signal structure for L1 ($154f_0$) and L2 ($120f_0$) is shown in Figure 4.5 (where f_0 is the fundamental frequency: 10.23 MHz). As shown in Figure 4.5, the L1 frequency ($154f_0$) is modulated by two PRN codes (plus the navigation message data), the C/A code, and the P code. The L2 frequency ($120f_0$) is modulated by only one PRN code at a time. One of the P code modes has no data modulation. The nominal reference frequency, f_0, as it appears to an observer on the ground, is 10.23 MHz. To compensate for relativistic effects, the output of the SV's frequency standard (as it appears from the SV) is 10.23 MHz offset by a $\Delta f/f$ of $4.467 \quad 10^{-10}$ (see Section 7.2.3). This results in a Δf of $4.57 \quad 10^{-3}$ Hz and $f_0 = 10.22999999543$ MHz [10]. To the GPS receiver on the ground, the C/A code has a chipping rate of $1.023 \quad 10^6$ chips/s ($f_0/10 = 1.023$ MHz) and the P code has a chipping rate of $10.23 \quad 10^6$ chips/s ($f_0 = 10.23$ MHz). Using the notation introduced in Section 4.2.3, the C/A code signal uses a BPSK-R(1) modulation and the P code uses a BPSK-R(10) modulation. The P code is available to PPS users but not to SPS users since the CS normally configures an AS mode in the SV. When AS is

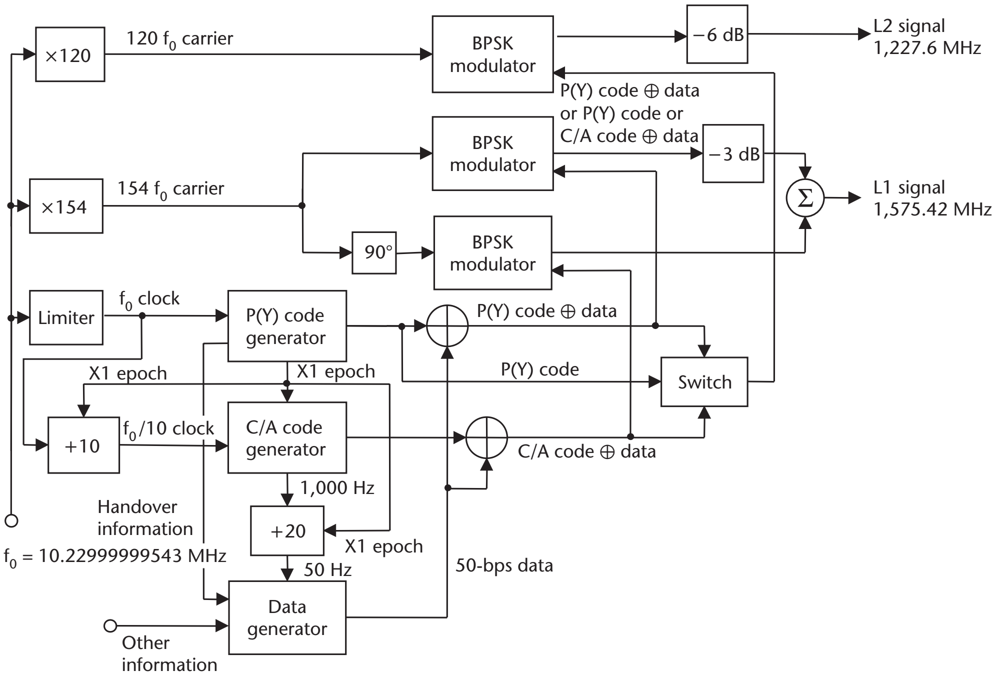


Figure 4.5 Legacy GPS satellite signal structure.

activated, the P code is encrypted to form what is known as the Y-code. The Y-code has the same chipping rate as the P code. Thus, the acronym often used for the precision (encrypted) code is P(Y) code.

Since the PPS (primarily military) users have access to the cryptographic keys and algorithms used in the AS process but the SPS (primarily civil) users do not, then AS denies access to the P code by SPS users. In the past, both the C/A code and the P(Y) code, as well as the L1 and L2 carrier frequencies, were subjected to an encrypted time-varying frequency offset (referred to as *dither*) plus an encrypted ephemeris and almanac offset error (referred to as *epsilon*) known as SA. SA denied the full accuracy of GPS to the stand-alone SPS users. However, SA has been deactivated on all GPS satellites since May 1, 2000, so this subject will not be further discussed in this chapter.

Note in Figure 4.5 that the same 50-bps navigation message data is combined with both the C/A code and the P(Y) code prior to modulation with the L1 carrier. An exclusive-or logic gate is used for this modulation process, denoted by ⊕. Since the C/A code ⊕ data and P(Y) code ⊕ data are both synchronous operations, the bit transition rate cannot exceed the chipping rate of the PRN codes. Also note that BPSK modulation is used with the carrier signals. The P(Y) code ⊕ data is modulated in phase quadrature with the C/A code ⊕ data on L1. As shown in Figure 4.5, the L1 carrier is phase shifted 90° before being BPSK modulated by the C/A code ⊕ data. Then this result is combined with the attenuated output of the BPSK modulation of L1 by the P(Y) code ⊕ data. The 3-dB amplitude difference and phase relationship between P code and C/A code on L1 are illustrated by the vector phase diagram in Figure 4.6. Figure 4.7 illustrates the result of P code ⊕ data and C/A ⊕ data. As observed in Figure 4.7, the exclusive-or process is equivalent to binary multiplica-

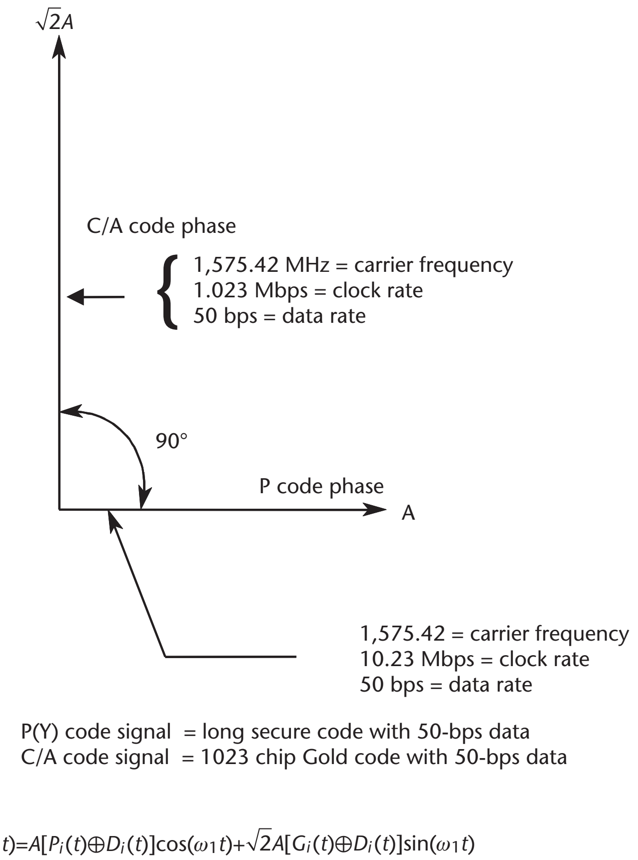


Figure 4.6 GPS signal structure for L1.

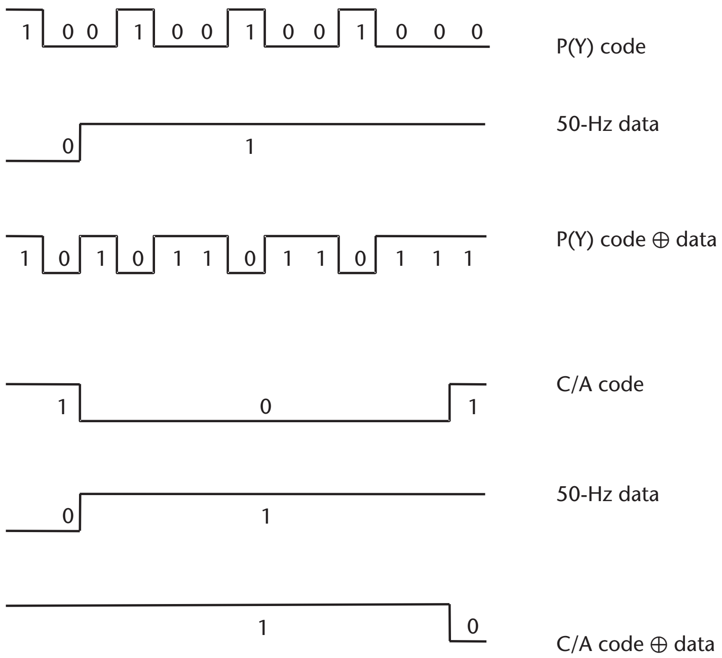


Figure 4.7 GPS code mixing with data.

tion of two 1-bit values yielding a 1-bit product using the convention that logical 0 is plus and logical 1 is minus. There are 204,600 P(Y) code epochs between data epochs and 20,460 C/A code epochs between data epochs, so the number of times that the phase could change in the PRN code sequences due to data modulation is relatively infrequent, but the spectrum changes due to this modulation are very significant.

Figure 4.8 illustrates how the signal waveforms would appear before and after the BPSK modulation of one P(Y) code ⊕ data transition and one C/A code ⊕ data transition. There are 154 carrier cycles per P(Y) code chip and 1,540 carrier cycles per C/A code chip on L1, so the phase shifts on the L1 carrier are relatively infrequent. The L2 frequency (1,227.60 MHz) can be modulated by either the P(Y) code ⊕ data or the C/A code ⊕ data or by the P(Y) code alone as selected by the CS. The P(Y) code and C/A codes are never present simultaneously on L2 prior to GPS modernization (see Section 4.5), unlike the case with L1. In general, the P(Y) code ⊕ data is the one selected by the CS. There are 120 carrier cycles per P(Y) code chip on L2, so the phase transitions on the L2 carrier are relatively infrequent. Table 4.2 summarizes the GPS signal structure on L1 and L2.

The PPS user has the algorithms, the special Y-code hardware per channel, and the key to gain access to the Y-code. PPS receivers formerly included a precise positioning service security module (PPSSM) to store and process the cryptographic keys and an auxiliary output chip (AOC) to produce the Y-code. Current generation PPS receivers are built around a security architecture referred to as the selective availabil-

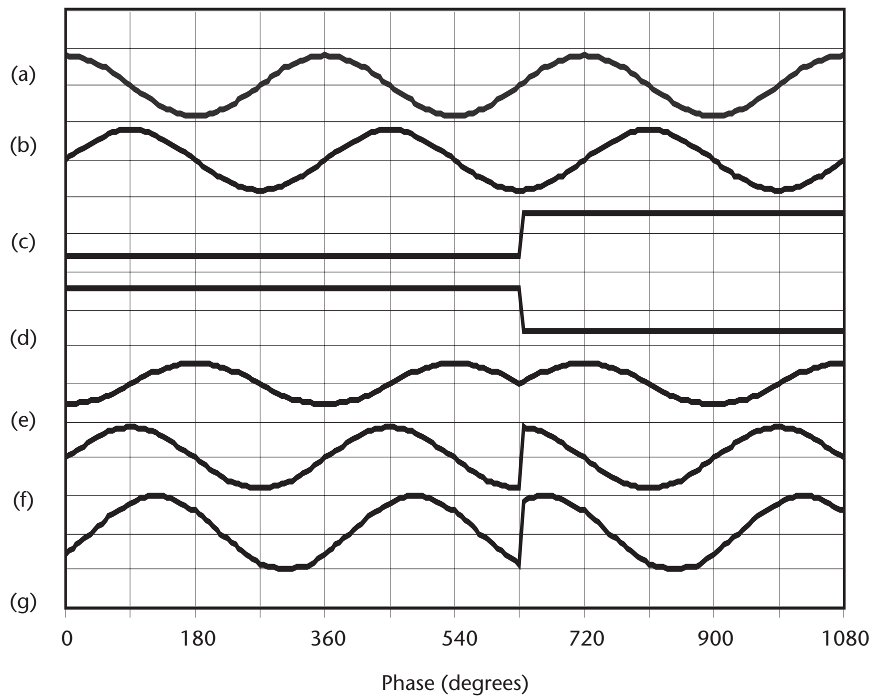


Figure 4.8 GPS L1 carrier modulation: (a) L1 carrier (0° phase), (b) L1 carrier (90° phase), (c) P(Y) code ⊕ data, (d) C/A code ⊕ data, (e) P(Y) code ⊕ data BPSK modulated on L1 carrier (0° phase) with 3-dB attenuation, (f) C/A code ⊕ data BPSK modulated on L1 carrier (90° phase), and (g) composite modulated L1 carrier signal.

Table 4.2 Legacy GPS Signal Structure

Signal Priority	*Primary*	*Secondary*
Signal designation	L1	L2
Carrier frequency (MHz)	1,575.42	1,227.60
PRN codes (Mchip/s)	P(Y) = 10.23 and C/A = 1.023	P(Y) = 10.23 or C/A = 1.023 (Note 1)
Navigation message data modulation (bps)	50	50 (Note 2)

1. The code usually selected by the CS on L2 is P(Y) code.
2. The 50-Hz navigation data message is usually modulated on L2 P(Y) code but can be turned off by the CS. There are three possibilities on L2: P(Y) code with data, P(Y) code with no data, and C/A code with data.

ity/antispoofing module (SAASM). The use of the AS Y-code denies direct (SPS GPS receiver) access to the precision code. This significantly reduces the possibility of an enemy spoofing a PPS receiver (i.e., transmitting a stronger, false precise code that captures and misleads the receiver). However, AS also denies direct access to the precision code to all SPS users, friendly or otherwise. Indirect access is still possible as discussed in [11] and Section 5.14.

4.4 Navigation Message Format

As described in Section 4.3, both the C/A code and P(Y) code signals are modulated with 50-bps data. This data provides the user with the information necessary to compute the precise locations of each visible satellite and time of transmission for each navigation signal. The data also includes a significant set of auxiliary information that may be used, for example, to assist the equipment in acquiring new satellites, to translate from GPS system time to UTC (see Section 2.6), and to correct for a number of errors that affect the range measurements. This section outlines the main features of the GPS navigation message format. For a more complete description, the interested reader is referred to [10].

Table 4.8 C/A Code Maximum Cross-Correlation Power (Zero Doppler Differences)

Cumulative Probability of Occurrence	*Cross-Correlation for Any Two Codes (dB)*
0.23	–23.9
0.50	–24.2
1.00	–60.2

Table 4.9 C/A Code Maximum Cross-Correlation Power Summed for All 32 Codes (Increments of 1-kHz Doppler Differences)

Cumulative Probability of Occurrence	*Cross-Correlation at Δ = 1 kHz (dB)*	*Cross-Correlation at Δ = 2 kHz (dB)*	*Cross-Correlation at Δ = 3 kHz (dB)*	*Cross-Correlation at Δ = 4 kHz (dB)*	*Cross-Correlation at Δ = 5 kHz (dB)*
0.001	–21.1	–21.1	–21.6	–21.1	–21.9
0.02	–24.2	–24.2	–24.2	–24.2	–24.2
0.1	–26.4	–26.4	–26.4	–26.4	–26.4
0.4	–30.4	–30.4	–30.4	–30.4	–30.4

The GPS navigation message is transmitted in five 300-bit subframes, as shown in Figure 4.19. Each subframe is itself composed of ten 30-bit words. The last 6 bits in each word of the navigation message are used for parity checking to provide the user equipment with a capability to detect bit errors during demodulation. A (32, 26) Hamming code is employed. The five subframes are transmitted in order beginning with subframe 1. Subframes 4 and 5 consist of 25 pages each, so that the first time through the five subframes, page 1 of subframes 4 and 5 are broadcast. In the next cycle through the five subframes, page 2 of subframes 4 and 5 are broadcast and so on.

Although there are provisions for a loss of ground contact, normally the control segment uploads critical navigation data elements once or twice per day per satellite. In this nominal mode of operation, the same critical navigation data elements (e.g., satellite ephemeris and clock correction data) are broadcast repeatedly over 2-hour time spans (except if an upload occurs during this interval). On 2-hour boundaries, each satellite switches to broadcasting a different set of these critical elements, which are stored in tables in the satellite's RAM. The control segment generates these message elements based upon its current estimates of each satellite's position and clock error and prediction algorithms on how these parameters will change over time.

The first two words of each subframe (bits 1–60) contain telemetry (TLM) data and a handover word (HOW). The TLM word is the first of the 10 words in each subframe and includes a fixed preamble, a fixed 8-bit pattern 10001011 that never changes. This pattern is included to assist the user equipment in locating the beginning of each subframe. Each TLM word also includes 14 bits of data that are only meaningful to authorized users. The HOW, so-named because it allows the user

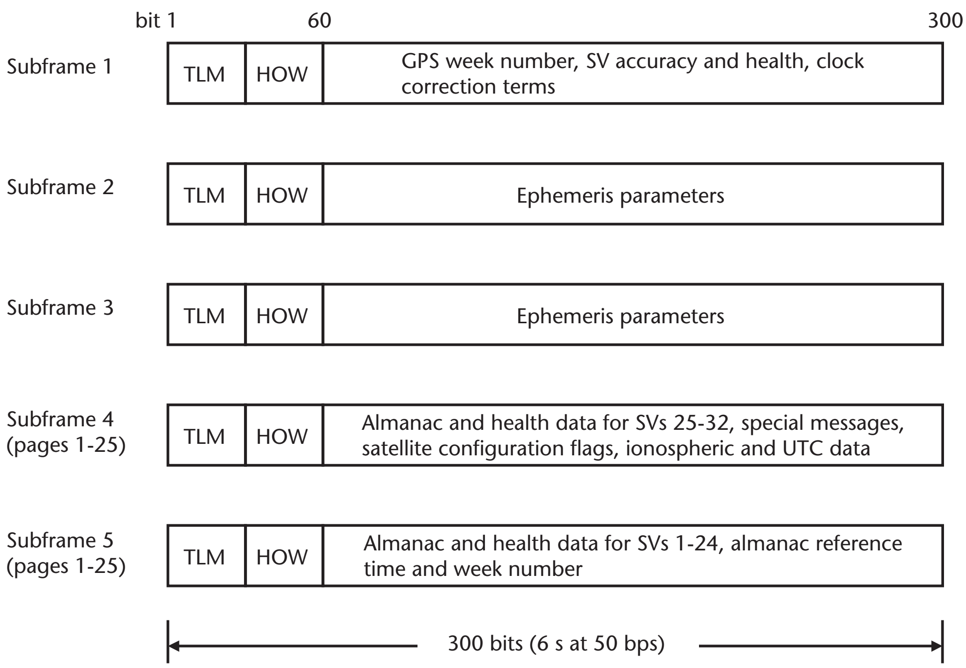


Figure 4.19 Navigation message format.

equipment to "handover" from C/A code tracking to P(Y) code tracking, provides the GPS time-of-week (TOW) modulo 6 seconds corresponding to the leading edge of the following subframe. The HOW also provides two flag bits, one that indicates whether antispoofing is activated (see Section 4.3.1), and one that serves as an alert indicator. If the alert flag is set, it indicates that the signal accuracy may be poor and should be processed at the user's own risk. Finally, the HOW provides the subframe number (1–5).

Subframe 1 provides the GPS transmission week number, which is the number of weeks modulo 1,024 that have elapsed since January 5, 1980. The first rollover of the GPS week number occurred on August 22, 1999. The next rollover will occur in April 2019. It is prudent that the GPS receiver designer keep track of these rare but inevitable rollover epochs in nonvolatile memory. Subframe 1 also provides the following satellite clock correction terms: a_{f0}, a_{f1}, a_{f2}, and time of clock, t_{oc}. These terms are extremely important for precise ranging, since they account for the lack of perfect synchronization between the timing of the SV broadcast signals and GPS system time (see Section 7.2.1). A 10-bit number referred to as issue of data, clock (IODC) is included in subframe 1 to uniquely identify the current set of navigation data. User equipment can monitor the IODC field to detect changes to the navigation data. The current IODC is different from IODCs used over the past seven days. Subframe 1 also includes a group delay correction, T_{gd}, a user range accuracy (URA) indicator, a SV health indicator, an L2 code indicator, and an L2 P data flag. T_{gd} is needed by single-frequency (L1- or L2-only) users since the clock correction parameters refer to the timing of the P(Y) code on L1 and L2, as apparent to a user that is using a linear combination of dual-frequency L1/L2 P(Y) code measurements to mitigate ionospheric errors (see Section 7.2.4.1). The URA indicator provides the user with an estimate of the 1-sigma range errors to the satellite due to satellite and control segment errors (and is fully applicable only for L1/L2 P-code users). The SV health indicator is a 6-bit field that indicates whether the satellite is operating normally or whether components of the signal or navigation data are suspected to be erroneous. The L2 code indicator field indicates whether the P(Y) code or C/A code is active on L2. Finally, the L2 P data flag indicates whether navigation data is being modulated onto the L2 P(Y) code.

Subframes 2 and 3 include the osculating Keplerian orbital elements described in Section 2.3 that allow the user equipment to precisely determine the location of the satellite. Subframe 2 also includes a fit interval flag and an age of data offset (AODO) term. The fit interval flag indicates whether the orbital elements are based upon a nominal 4-hour curve fit (that corresponds to the 2-hour nominal data transmission interval described earlier) or a longer interval. The AODO term provides an indication of the age of the elements based on a navigation message correction table (NMCT) that has been included in the GPS navigation data since 1995 [15]. Both subframes 2 and 3 also include an issue of data ephemeris (IODE) field. IODE consists of the 8 least significant bits (LSBs) of IODC and may be used by the user equipment to detect changes in the broadcast orbital elements.

Pages 2–5 and 7–10 of subframe 4 and pages 1–24 of subframe 5 contain almanac data (coarse orbital elements that allow the user equipment to determine approximate positions of other satellites to assist acquisition) for SVs 1–32 (see Section 2.3). Page 13 of subframe 4 includes the NMCT range corrections. Page 18 of subframe 4 includes ionospheric correction parameters for single-frequency users (see Section 7.1.2.5) and parameters so that user equipment can relate UTC to GPS system time (see Section 2.6.3). Page 25 of subframes 4 and 5 provide configuration and health flags for SVs 1–32. The data payloads of the remaining pages of subframes 4 and 5 are currently reserved.