CHAPTER 12

SATELLITE NAVIGATION AND THE GLOBAL POSITIONING SYSTEM

12.1 INTRODUCTION

The Global Positioning Satellite System (GPS) has revolutionized navigation and position location. It is now the primary means of navigation for most ships and aircraft and is widely used in surveying and many other applications. The GPS system, originally called NAVSTAR, was developed as a military navigation system for guiding missiles, ships, and aircraft to their targets. GPS satellites transmit L-band signals that are modulated by several codes. The C/A (*coarse acquisition*) code was made available to the public in the mid-1980s. The secure high accuracy P code allows authorized users (mainly military) to achieve positioning accuracy of 3 m. This was the accuracy that the military users wanted for targeting smart bombs and cruise missiles, but such accuracies are also useful for auto-landing aircraft in fog and for docking ships in bad weather.

The first commercial use of GPS was in surveying, but by 1990 several companies had produced low-cost, handheld GPS receivers for general position location and navigation. Increased sales and larger volume production quickly brought down the price of a GPS receiver, and the market expanded rapidly. GPS receivers are now a consumer product, and will soon be found in every car and cellular telephone.

The GPS system has been successful because it provides a direct readout of the present position of a GPS receiver with a typical accuracy of 30 m. There are other position location systems, such as LORAN, (a contraction of long range navigation) that can also provide direct readout of position, but not with the accuracy and reliability of GPS. The success of GPS is an excellent example of what satellites do best: *broadcasting*. An unlimited number of GPS receivers can operate simultaneously because all that a GPS receiver has to do to locate itself is to receive signals from four GPS satellites.

The GPS space segment consists of 24 satellites in medium earth orbit (MEO) at a nominal altitude of 20,200 km with an orbital inclination of 55°. The satellites are clustered in groups of four, called constellations, with each constellation separated by 60° in longitude. The orbital period is approximately one-half a sidereal day (11 h 58 min) so the same satellites appear in the same position in the sky twice each day. The satellites carry station-keeping fuel and are maintained in the required orbits by occasional station-keeping maneuvers, just like GEO satellites. The orbits of the 24 GPS satellites ensure that at any time, anywhere in the world, a GPS receiver can pick up signals from at least four satellites. Up to 10 satellites may be visible at some times, and more than four satellites are visible nearly all of the time. Replacement satellites are launched as needed, so there may be more than 24 operational GPS satellites at any given time.

Figure 12.1 shows a GPS satellite. The satellites weigh 1877 kg at launch and have a design lifetime of 10 years. In 2000, there were 30 GPS satellites in orbit, some of which were spares. Because GPS is an integral part of the defense of the United States, spare GPS satellites are kept in orbit and more spares are ready for immediate launch. The GPS

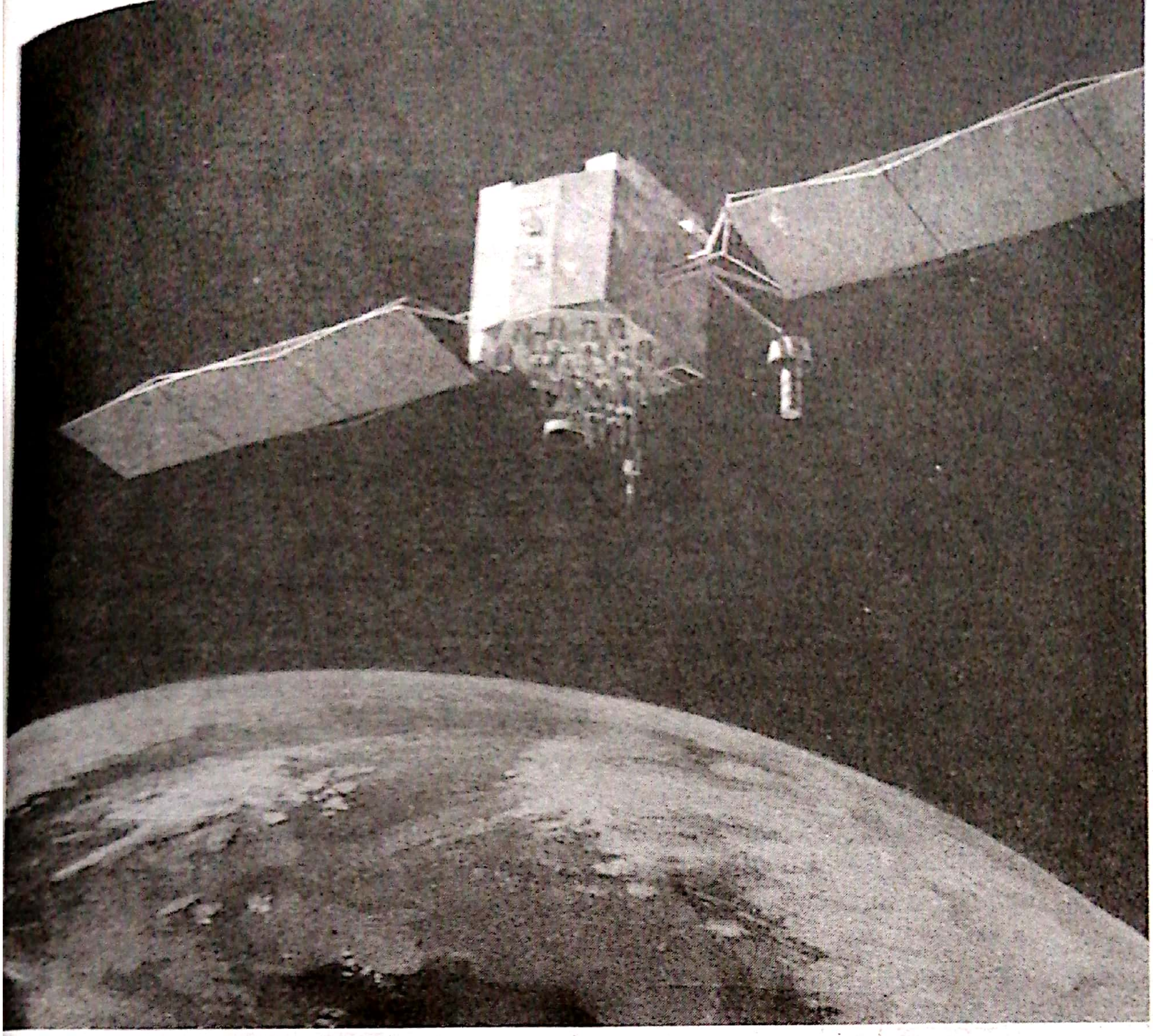

FIGURE 12.1 GPS block II-F satellite.

system is operated by the U.S. Air Force from the GPS master control station (MCS) at Falcon Air Force Base in Colorado Springs, Colorado. The MCS and a series of subsidiary control stations around the globe continuously monitor all GPS satellites as they come into view and determine the orbit of each satellite. The MCS and other stations calculate ephemeris data for each satellite, atomic clock error, and numerous other parameters needed for the *navigation message*. The data are then transmitted to the satellite using a secure S-band link and used to update onboard data. There are five GPS monitor stations located in Hawaii, Colorado Springs, Ascension Island in the Atlantic Ocean, Diego Garcia in the Indian Ocean and Kwajalein in the Pacific Ocean[1]. The monitor stations have precise cesium time standards and make continuous measurements of range to all visible satellites. These measurements are performed every 1.5 s, and used to provide updates for the navigation messages.

The position of a GPS receiver is found by trilateration, which is one of the simplest and most accurate methods of locating an unknown position. In trilateration, the distance of the unknown point from three known points is measured. The intersection of the arcs corresponding to three distances defines the unknown point relative to the known points, since three measurements can be used to solve three equations to give the latitude, longitude, and elevation of the receiver. The distance between a transmitter and a receiver can be found by measuring the time it takes for a pulse of RF energy to travel between the two. The distance is calculated using the velocity of electromagnetic waves in free space, which is assumed to be equal the velocity of light, c, with $c = 299{,}792{,}458$ m/s. Time can be measured electronically more accurately than any other parameter by the use of atomic clocks, and this is how the GPS position location system can achieve a measurement accuracy of 1 m in a distance of 20,000 km. To achieve a position location accuracy of 1 m, timing measurements must have an accuracy better than 3 ns. This is possible with modern digital circuitry and a great deal of averaging.

Each satellite carries several high accuracy atomic clocks and radiates a sequence of bits that starts at a precisely known time. A GPS receiver contains a clock that is synchronized in turn to the clock on each satellite that it is receiving. The receiver measures the time delay of the arrival of the bit sequence, which is proportional to the distance between the satellite and the GPS receiver. When the distance of a GPS receiver from three satellites has been measured, the remaining piece of information that is required is the position of each satellite. This is calculated in the GPS receiver using the ephemeris for the satellite orbits that are broadcast by each satellite in its navigation message. Since the time at which the transmitted bit sequence started is known at the receiver, the position of the satellite at that time can be calculated from its orbital data. Making the calculation for four satellites provides the receiver with sufficient information to determine its position with very good accuracy. Four satellites, rather than three, are needed because the clock in the receiver is not inherently accurate enough. The fourth distance measurement provides information from which clock errors in the receiver can be corrected and the receiver clock synchronized to GPS time with an accuracy better than 100 ns.

GPS satellites transmit two signals at different frequencies, known as L1 and L2. The L2 signal is modulated with a 10.23 Mbps pseudorandom (PN) bit sequence called the *P code* that is used by military positioning systems. The P code is transmitted in an encrypted form known as the *Y code*, which restricts the use of the P code to authorized users.

The L1 frequency carrier is modulated by a 1.023 Mbps PN sequence called the *C/A code* that is available for public use, and also carries the P code as a quadrature modulation. The higher bit rate of the P code provides better measurement accuracy than the 1.023 Mbps C/A code. C/A stands for *coarse acquisition* and P stands for *precise*. GPS systems using the secure Y code require the C/A code as an intermediate step in making distance measurements with high accuracy. The accuracy of C/A code receivers was deliberately degraded some of the time by a process called *selective availability* (SA). SA causes variations in the C/A code satellite transmissions that result in less accurate calculation of position. SA was discontinued in May 2000, but can be reinstituted if the President of the United States declares a National Emergency.

The GPS system provides two categories of service. The precise positioning service (PPS) receivers track both P code and C/A code on L1 and L2 frequencies. The PPS is used mainly by military users, since the P code is encrypted into the Y code before transmission and requires decryption equipment in the receiver. Standard positioning service (SPS) receivers track the C/A code on L1. This is the service that is used by the general public. The P(Y) and C/A codes transmitted by each satellite create direct sequence spread spectrum signals which occupy the same frequency bands. Both the C/A codes and the P codes are publicly available, but the P code cannot be recovered in a GPS receiver without a knowledge of the Y code decryption algorithm. In this discussion we will concentrate on the C/A code and its use in position location.

The former USSR built and operated a global navigation system that is very similar to GPS, known in the West as Glonass for global navigation satellite system. Almost everything about Glonass is similar to GPS except the multiple access technique. Glonass uses FDMA, with a different transmit frequency at each satellite. The equivalents of the P code and C/A code can be transmitted by Glonass satellites in RF bandwidths of 20 kHz and 2 kHz, so 100 satellites can be accommodated in a bandwidth of 2 MHz. An FDMA receiver with 100 channels is simpler than a CDMA receiver. A frequency synthesizer that can be tuned to the unique frequency of each satellite is required, rather than the digital correlators that recover the GPS signals in a CDMA receiver.

The European Union is considering building a similar satellite navigation system called Galileo, scheduled for operation by 2008, to provide precise navigation signals without dependence on the United States.

12.2 RADIO AND SATELLITE NAVIGATION

Prior to the development of radio, navigation was by compass and landmarks on land, and by the sun and stars at sea. Neither technique provides high accuracy, and shipwrecks caused by inaccurate navigation and foggy weather were a common occurrence. On land, people often got lost in wilderness areas (and still do). Pilots of light aircraft, relying solely on a map and landmarks, would get lost and run out of fuel before they found somewhere to land. With a GPS receiver and a map, it is impossible to get lost. GPS receivers are very popular with airplane pilots, owners of sea-going boats, and wilderness hikers.

The development of aircraft that could fly above the clouds, and particularly the building of large numbers of bomber aircraft in the 1930s, made radio navigation essential. Military thinking after WWI, and during WWII, placed high reliance on the ability of bomber aircraft to win a war by destroying the weapon manufacturing capability of the enemy. During WWII, the allies sent 1000 bomber aircraft at a time to targets in Germany, causing immense destruction to many cities. The philosophy of mass destruction continued after WWII with the development of nuclear bombs, intercontinental ballistic missiles (ICBMs), and cruise missiles. However, bomber aircraft, ICBMs, and cruise missiles must find their targets, so accurate navigation is an essential part of each of these weapon systems. This demand for accurate targeting of airborne weapons led to the development of GPS. However, the majority of GPS users are now civilian, and the worldwide market for GPS equipment is projected to be worth \$25 B by 2005.

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

GPS can provide a single navigation system with better accuracy and reliability than all earlier radio navigation aids. It can provide navigation of aircraft directly between airports, instead of indirectly via airways, while providing absolute position readout of latitude and longitude. Differential GPS can be used instead of ILS to provide the required straight line in the sky for an instrument approach to a runway, and can be linked to an autopilot to provide automatic landing of aircraft in zero visibility conditions. Ships can safely navigate and dock in treacherous waters in bad weather by using differential GPS. Eventually, GPS will replace all other means of navigation, although some may be retained as backup systems in case of failure of the GPS receiver(s) or jamming of the signals.

GPS was preceded by an earlier satellite navigation system called *Transit*, built for the U.S. Navy for ship navigation, which achieved much lower accuracy and became obsolete when GPS was introduced. Transit satellites were in low earth orbits and the system used the Doppler shift observed at the receiver when a beacon signal was transmitted by the satellite. Because of the high velocity of LEO satellites—about 7.5 km/s—their signals are significantly shifted up in frequency when the satellite appears over the horizon with a component of velocity toward the receiver. The Doppler shift falls to zero as the satellite passes the observer, and then becomes negative as the satellite flies away.

Observation of the Doppler shift with time, which may need to be as long as 10 min, and a knowledge of the satellite orbit, allows calculation of the receiver's position.

There was never a sufficient number of Transit satellites to provide continuous position data, and the long time required to obtain an accurate position fix was a disadvantage. A similar system called *SARSAT*, for search and rescue satellite, is used to find emergency locator transmitters (ELTs) on aircraft that have crashed. Most general aviation aircraft carry an ELT, which turns on at a frequency of 121.5 MHz when subjected to high G forces, as might be experienced if the aircraft crashes. Certain LEO satellites carry 121.5-MHz receivers that relay the signals to earth stations at rescue coordination centers. If an aircraft ELT turns on, a SARSAT satellite will eventually fly by and relay a Doppler shifted signal to the rescue station.

Analysis of the Doppler shift over the observation period provides information about the location of the ELT, but with an accuracy of only 1 or 2 km. Almost 97% of ELT locations turn out to be false alarms—the ELT was dropped or accidentally turned on. It seems probable that GPS and cellular phones or satellite phones will eventually replace the SARSAT system.

12.3 GPS POSITION LOCATION PRINCIPLES

The basic requirement of a satellite navigation system like GPS is that there must be four satellites transmitting suitably coded signals from known positions. Three satellites are required to provide the three distance measurements, and the fourth to remove receiver clock error. Figure 12.2 shows the general arrangement of position location with GPS. The three satellites provide distance information when the GPS receiver makes three measurements of *range*, R_i, from the receiver to three known points. Each distance R_i can be thought of as the radius of a sphere with a GPS satellite at its center. The receiver lies at the intersection of three such spheres, with a satellite at the center of each sphere. Locally, at the receiver, the spheres will appear to be planes since the radii of the spheres are very large. A basic principle of geometry is that the intersection of three planes completely defines a point. Thus three satellites, through measurement of their distances to the receiver, define the receiver location close to the earth's surface. There is another point in outer space where the three spheres intersect, but it is easily eliminated in the calculation process.

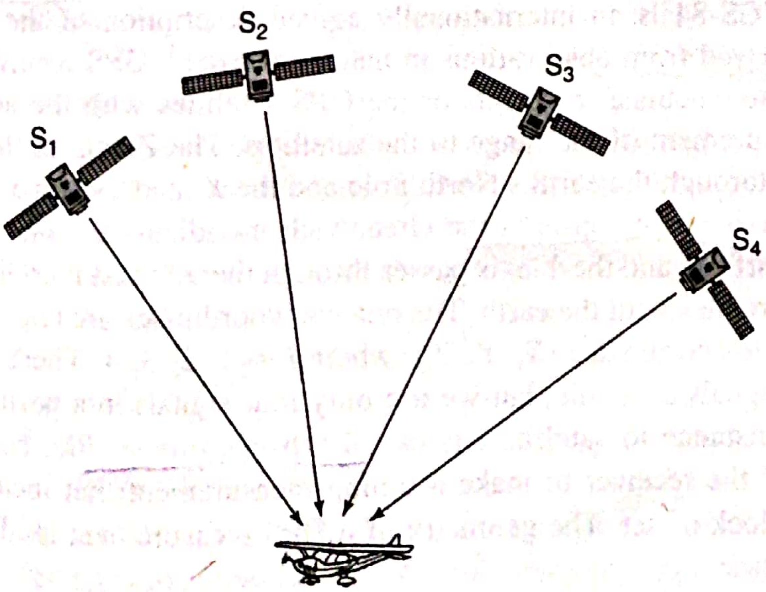

FIGURE 12.2 General arrangement of position locations with GPS. The aircraft must receive signals from four GPS satellites to be able to determine its position.

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

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

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

Position Location in GPS

First, we will define the coordinates of the GPS receiver and the GPS satellites in a rectangular coordinate system with its origin at the center of the earth. This is called the earth centered earth fixed (ECEF) coordinate system, and is part of the WGS-84 description of the earth. WGS-84 is an internationally agreed description of the earth's shape and parameters, derived from observations in many countries[4]. GPS receivers use the WGS-84 parameters to calculate the orbits of the GPS satellites with the accuracy required for precise measurement of the range to the satellites. The Z-axis of the coordinate system is directed through the earth's North Pole and the X- and Y-axes are in the equatorial plane. The X-axis passes through the Greenwich meridian—the line of zero longitude on the earth's surface, and the Y-axis passes through the 90° east meridian. The ECEF coordinate system rotates with the earth. The receiver coordinates are (U_x, U_y, U_z), and the four satellites have coordinates (X_i, Y_i, Z_i), where $i = 1, 2, 3, 4$. There may be more than four satellite signals available, but we use only four signals in a position calculation. The measured distance to satellite i is called a *psuedorange*, PR_i, because it uses the internal clock of the receiver to make a timing measurement that includes errors caused by receiver clock offset. The geometry of a GPS measurement is illustrated in Figure 12.3.

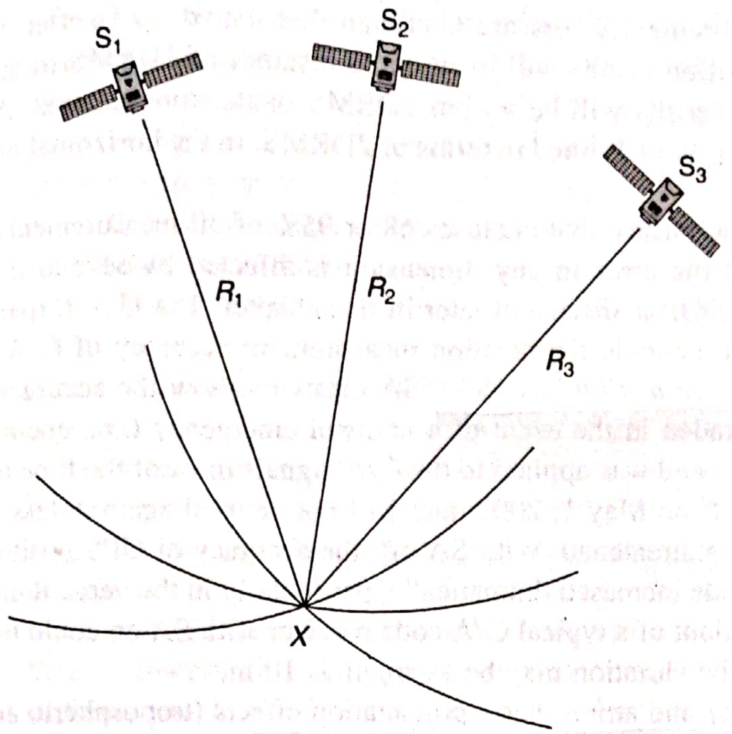

FIGURE 12.3 Position location by measurement of the distance to three satellites. The GPS receiver is located at point X, where three spheres with radii R_1, R_2, and R_3 intersect. The centers of the spheres are the three GPS satellites S_1, S_2, and S_3. If the distances R_1, R_2, and R_3 are measured, the location of the point X can be uniquely defined.

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

$$PR_i = T_i \times c \tag{12.1}$$

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

$$R^2 = (x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 \tag{12.2}$$

The equations which relate pseudorange to time delay are called *ranging equations*:

$$\begin{aligned}(X_1 - U_x)^2 + (Y_1 - U_y)^2 + (Z_1 - U_z)^2 &= (PR_1 - \tau c)^2\\(X_2 - U_x)^2 + (Y_2 - U_y)^2 + (Z_2 - U_z)^2 &= (PR_2 - \tau c)^2\\(X_3 - U_x)^2 + (Y_3 - U_y)^2 + (Z_3 - U_z)^2 &= (PR_3 - \tau c)^2\\(X_4 - U_x)^2 + (Y_4 - U_y)^2 + (Z_4 - U_z)^2 &= (PR_4 - \tau c)^2\end{aligned} \tag{12.3}$$

where τ is receiver clock error (offset, or *bias*).

The position of the satellite at the instant it sent the timing signal (which is actually the start of a long sequence of bits) is obtained from ephemeris data transmitted along with the timing signals. Each satellite sends out a data stream that includes ephemeris data for itself and the adjacent satellites. The receiver calculates the coordinates of the satellite relative to the center of the earth, (X_i, Y_i, Z_i), and then solves the four ranging equations for the four unknowns using standard numerical techniques for the solution of nonlinear simultaneous equations. (The equations are nonlinear because of the squared terms.)

The four unknowns are the location of the GPS receiver, (U_x, U_y, U_z), relative to the center of the earth and the clock offset τ—called *clock bias* in GPS terminology. The receiver position is then referenced to the surface of the earth, and can be displayed in latitude, longitude, and elevation. Typical accuracy for a low-cost GPS receiver using the GPS C/A code is 30 m defined as a 2DRMS error. The term DRMS means the distance root mean square error of the measured position relative to the true position of the receiver. If the measurement errors are Gaussian distributed, as is often the case, 68% of the measured position results will be within a distance of 1DRMS from the true location and 95% of the results will be within 2DRMS of the true location. Accuracy in GPS measurements is usually defined in terms of 2DRMS, in the horizontal or vertical plane.

In practice, the error surface that encloses 68 or 95% of all measurements is not a circle but an ellipse, and the error in any dimension is affected by several *dilution of precision* (DOP) factors. DOP is discussed later in this chapter. The U.S. Department of Defense has the ability to degrade the position measurement accuracy of C/A code receivers by applying *selective availability* (SA). SA exists to allow the accuracy of C/A code receivers to be degraded in the event of a national emergency (i.e., enemy action) affecting the United States and was applied to the C/A signals most of the time until May 2000. SA was switched off on May 1, 2000, and will not be used again unless the security of the United States is threatened. With SA off, the accuracy of GPS position measurements with the C/A code increased dramatically, particularly in the vertical dimension. Variation in elevation readout of a typical C/A code receiver with SA on could be as large as 200 m. With SA off, the variation may be as small as 10 m.

Selective availability and atmospheric propagation effects (tropospheric and ionospheric) all cause errors in the timing measurements made by a GPS receiver, leading to position location errors. The atmosphere and the ionosphere introduce timing errors because the propagation velocity of the GPS signals deviates from the assumed free space value. The errors can be largely removed if a number of GPS reference stations are built at precisely known locations. The stations observe the GPS signals and compute the current error in position as calculated from GPS data. This information can then be broadcast to all GPS users as a set of corrections to be applied to GPS measurements. The system is called a *wide area augmentation system* (WAAS).

A network of 24 WAAS stations built in North America for the U.S. Federal Aviation Administration (FAA) provides aircraft with improved position measurement accuracy. Using WAAS, accuracies of a few meters can be obtained with C/A code receivers. In the event of a national emergency, WAAS would be switched off to prevent enemies using GPS for accurate targeting of weapons. WAAS also includes an integrity monitoring system to ensure that the GPS signals used by aircraft do not contain errors which could cause false readings. WAAS is required to send a warning of possible errors within 5.6 s if a problem is detected with any GPS satellite signal.

Similarly, a single reference station at a known location—for example, an airport—can determine the local measurement error in GPS and broadcast this information to GPS users so that greater accuracy can be obtained with a C/A code receiver. This is one (simple) form of *differential GPS* (DGPS). More complex forms of differential GPS use a reference station which transmits the signals received from each GPS satellite so that phase comparisons can be made by the receiver. With lengthy integration times and a sophisticated phase comparison receiver, differential GPS accuracies of 1 cm can be obtained. With DGPS, the receiver computes its position relative to the reference station rather than in latitude and longitude. Differential GPS is used when a vehicle needs to be positioned accurately with respect to a fixed point, such as an aircraft with respect to a runway or a ship with respect to a berth.

GPS Time

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

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

Every GPS receiver is automatically synchronized to every other GPS receiver anywhere in the world through GPS time. This makes every GPS receiver a super clock, which knows time more accurately than any other time standard. Prior to the widespread use of GPS receivers, standard time transmissions were broadcast by the U.S. National Institute of Science and Technology (NIST, formerly the Bureau of Standards). The broadcasts were made in the HF (shortwave) band, and could be received throughout the United States. However, the HF signals propagate over long distance by reflection from the ionosphere, which introduces an uncertain delay into the time of arrival of the signal. The time standard provided by GPS is typically accurate to better than 170 ns, and has been used to synchronize electric power generators across the United States, for scientific applications that require synchronized clocks in different locations, and as a long-term frequency standard.

The time standard on board each GPS satellite consists of two cesium clocks plus two rubidium clocks (atomic clocks). An atomic clock uses the fundamental resonance of the cesium or rubidium molecule as a frequency reference to lock a crystal oscillator. In the GPS satellites, the master oscillator is at 10.23 MHz; all code rates, the L1, and the L2 RF frequencies are multiples or submultiples of 10.23 MHz. The atomic clocks are updated by the controlling ground stations to keep them within 1 μs of Universal Time Coordinated (UTC), and the navigation message broadcast by each satellite contains information about its current clock errors relative to GPS time. (UTC is a worldwide time standard. Greenwich Mean Time (GMT) is equal to UTC.)

12.4 GPS RECEIVERS AND CODES

GPS satellites transmit using *pseudorandom sequence* (PN) codes. All satellites transmit a C/A code at the same carrier frequency, 1575.42 MHz, called *L1*, using BPSK modulation. The L1 frequency is 154 times the master clock frequency of 10.23 MHz. The C/A code has a clock rate of 1.023 MHz and the C/A code sequence has 1023 bits, so the PN sequence lasts exactly 1.0 ms. The exact values of the frequencies are about 0.005 Hz lower than stated here to allow for relativistic effects caused by the high velocity of the satellites in their orbits (3.865 km/s). (GPS measurements are one of the few examples where relativistic effects must be taken into account, because the clocks are mounted on platforms moving at very high speeds.)

The P code is transmitted using BPSK modulation at the L2 carrier frequency of 1227.6 MHz (120 × 10.23 MHz), and is also transmitted with BPSK modulation on the L1 carrier frequency, in phase quadrature with the C/A code BPSK modulation. Figure 12.4 shows the way in which the L1 and L2 signals are generated on board a GPS satellite.

The C/A and P code transmissions from all GPS satellites are overlaid in the L1 and L2 frequency bands, making GPS a direct sequence spread spectrum (DS-SS) system

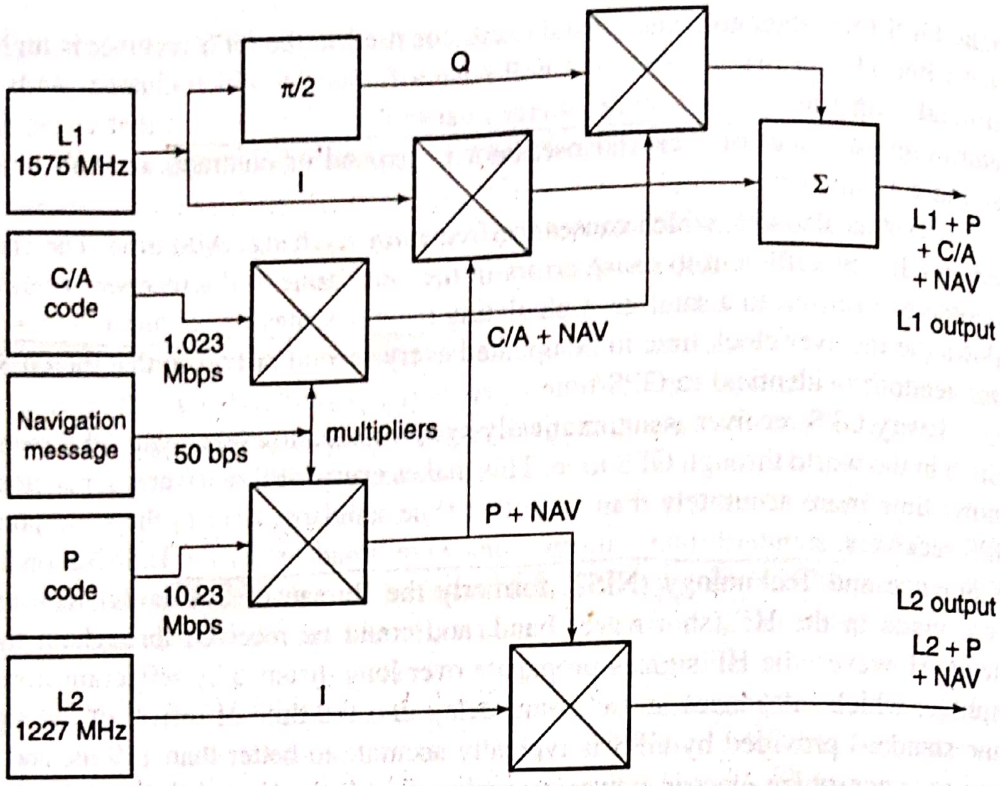

FIGURE 12.4 Signal generation in a GPS satellite.

(see Chapter 6 for details of spread spectrum techniques). The receiver separates signals from individual GPS satellites using knowledge of the unique C/A code that is allocated to each satellite. At most, 12 GPS satellites can be seen by a receiver at any one time, so the coding gain in the spread spectrum receiver must be sufficient to overcome the interference created by 11 unwanted signals while recovering the twelfth wanted signal.

The C/A Code

The C/A codes transmitted by GPS satellites are all 1023 bit *Gold codes*. GPS C/A Gold codes are formed from two 1023 bit *m-sequences*, called G1 and G2, by multiplying together the G1 and G2 sequences with different time offsets. An m-sequence is a maximum length pseudorandom (PN) sequence, which is easy to generate with a shift register and feedback taps. A shift register with n stages can generate a PN sequence $2^n - 1$ bits in length. The bit pattern is set by the feedback taps and combining logic of the shift register. The PN sequences G1 and G2 are both generated by 10-bit shift registers and are therefore both 1023 bits long. The clock rate for the C/A code is 1.023 MHz, so each sequence lasts 1.0 ms. Figure 12.5 shows a generator diagram for the C/A code.

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

$$C_i(t) = G1(t) \times G2(t + 10iT_c) \tag{12.4}$$

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

There are 64 Gold sequences available for satellites numbered 1 through 64. A total of 100 Gold sequences can be created using the algorithm in Eq. (12.4), but not all the

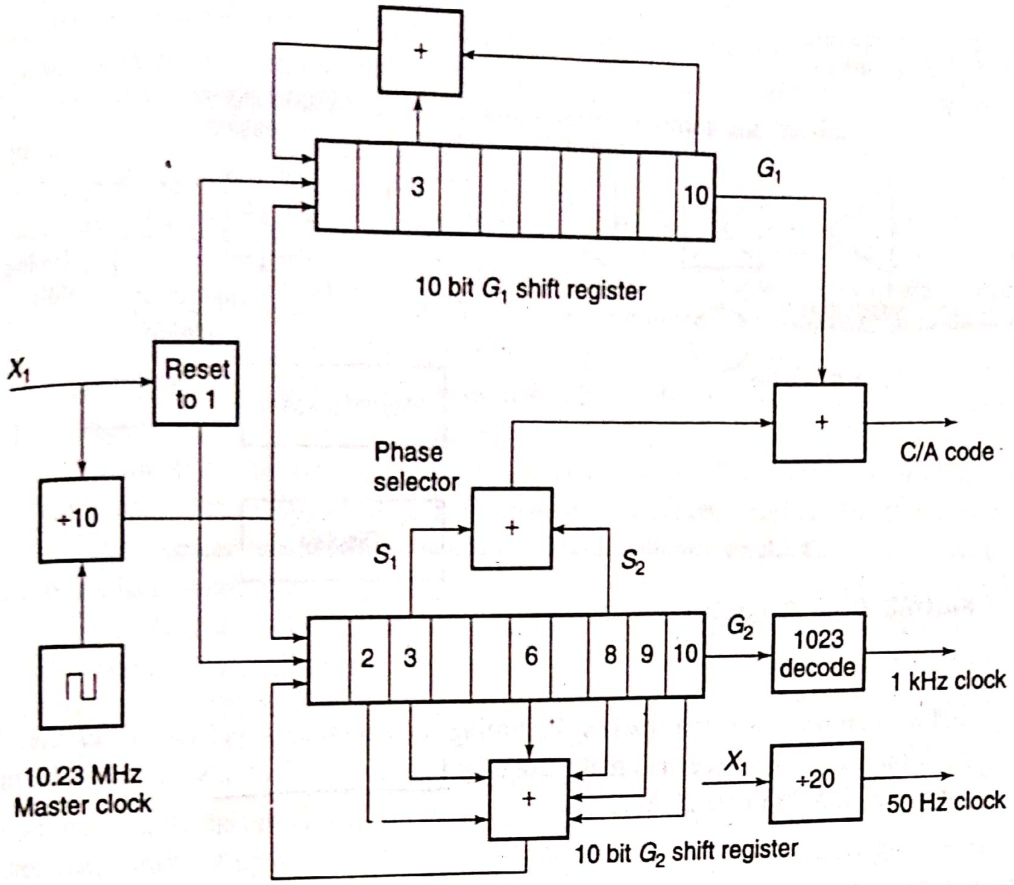

FIGURE 12.5 C/A code generator.

sequences have sufficiently low cross-correlation properties, and reference 4 states that only 37 are actually used in the GPS system. Low cross-correlation of the sequences is a requirement because the GPS receiver can pick up signals from as many as 12 satellites at the same time.

A correlator in the receiver looks for one of the sequences and must reject all other sequences that are present. Two C/A code sequences with zero cross-correlation would achieve a rejection ratio of 1023, but the 64 available C/A code sequences will not all have zero cross-correlation. The selected group of 37 are the sequences with the lowest levels of cross-correlation among the available set of 100 Gold code sequences. They also have low autocorrelation time sidelobes, another requirement of direct sequence spread spectrum systems.

The C/A code sequence length of 1.000 ms gives range ambiguity of 300 km, since the code travels at a velocity of approximately 3×10^8 m/s and therefore has a length in space of 3×10^5 m. The entire C/A code sequence repeats in space every 300 km, leading to ambiguity of position only if the GPS receiver is in outer space. The ambiguity is easily resolved if the receiver knows roughly where it is; just knowing that the receiver is located close to the earth's surface is usually sufficient. The user can enter the approximate location of the GPS receiver when it is first switched on to help resolve any ambiguities quickly.

Figure 12.6 shows a simplified block diagram of a C/A code GPS receiver. The antenna is typically a circularly polarized patch antenna with an LNA mounted on the printed circuit board. A conventional superhet receiver is used to generate an IF signal in a bandwidth of about 2 MHz, which is sampled using I and Q sampling techniques and processed digitally. The digital portion of the receiver includes a C/A code generator, a correlator,

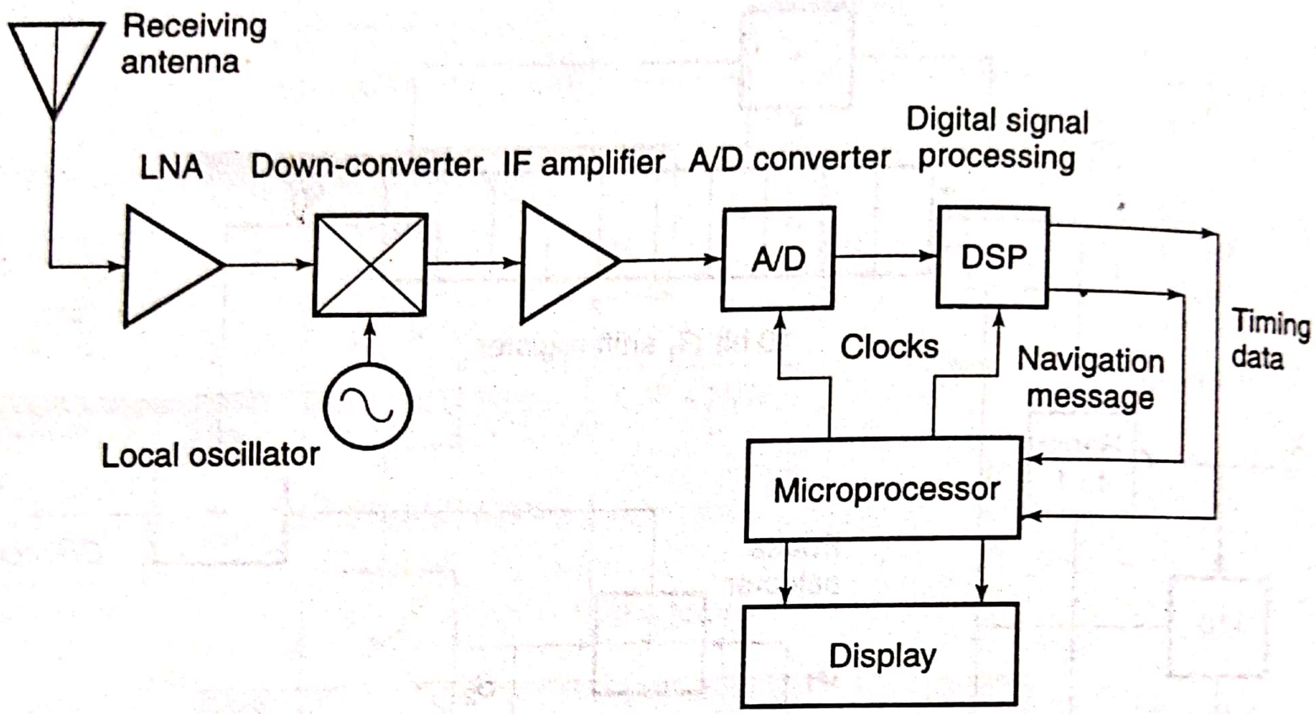

FIGURE 12.6 Simplified GPS receiver.

and a microprocessor that makes the timing measurements and calculates the receiver's position. Most GPS receivers make use of a 12-channel IC chip set that can be purchased for about $25.00 (Year 2000 prices).

12.10 GPS C/A CODE ACCURACY

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

- Satellite clock and ephemeris errors
- Selective availability (when switched on)
- Ionospheric delay (advance)
- Tropospheric delay
- Receiver noise
- Multipath

The accuracy that can be achieved with a GPS C/A code receiver can be found by using a range error budget. The figures in square brackets [ ] are for the case when selective availability (SA) is turned off. Typical values of range error are given in Table 12.4. All values are in meters (m). Note that a value of 2.4 m error is assigned to receiver noise. The value calculated in Section 12.8 was 4.2 m, for a worst-case received signal strength.

The range error introduced by the ionosphere and the troposphere can be partially removed by receiving identical signals at two different carrier frequencies. This technique is used by high precision P code receivers. The P code signal is transmitted on the L1 carrier at 1575.42 MHz, in phase quadrature with the C/A code signal. The P code is also transmitted on the L2 carrier at 1227.60 MHz. Algorithms are used in the P code receiver to calculate the net delay of the signal caused by the ionosphere and the atmosphere, and to then remove the errors from the calculated ranges. C/A code receivers use a standard atmosphere and ionosphere and assume a constant delay at a given elevation angle. Variations in the density of the atmosphere with atmospheric pressure changes, and in the free

TABLE 12.4 Range Error for C/A Code Measurements (m)

Satellite clock error	3.5	
Ephemeris errors	4.3	
Selective availability	32.0	[0]
Ionospheric delay	6.4	
Tropospheric delay	2.0	
Receiver noise	2.4	
Multipath	3.0	
RMS range error	33.4 m with SA	
[RMS range error	9.5 m without SA]	

Brackets indicate SA off.

electron content of the ionosphere, lead to departure from the standard values and hence to errors in the pseudorange calculation. There are plans to transmit the C/A code at a third and a fourth L-band frequency from later GPS satellites to provide improved accuracy with C/A code receivers[8].

The range error shown in Table 12.4 is for one satellite–earth path, for the pseudorange that is calculated from the timing measurements using the receiver clock. However, four pseudorange measurements are needed to make a position determination. Thus the position location output of the GPS receiver combines four path errors, which are not necessarily equal because of the geometry of the satellites in the sky and the different signal strengths at the receiver input. Receiver position is calculated in (x, y, z) coordinates, and the errors in x, y, and z depend on the elevation angle of satellites, the satellite geometry, and the other parameters in the error budget. The calculated position will have different levels of error in the x, y, and z directions. To account for these differences several *dilution of precision factors* (DOP) are defined. A DOP factor multiplies the basic position measurement error to give a larger error caused by the particular DOP effect.

Dilution of Precision: HDOP, VDOP, and GDOP

Horizontal dilution of precision is one of the most important DOP factors for most GPS users. It provides an error metric for the x and y directions, in the horizontal plane. A typical HDOP value is 1.5, and it is often the smallest of the DOPs. Horizontal measurement error for a C/A code receiver is typically 14.3 m with SA off (1DRMS) and 50 m with SA on (1DRMS). GPS practice uses 2DRMS as the quantifier for accuracy in position determination giving a 2DRMS accuracy of 28.6 m with SA off. The 2DRMS accuracy figure means that 95% of all measurements yield a position within 28.6 m of the true location of the GPS receiver, in this example.

There are many DOP factors in GPS. The more important ones are horizontal dilution of precision, HDOP, vertical dilution of precision, VDOP, and geometric dilution of precision, GDOP. Other DOPs include position dilution of precision, PDOP, and time dilution of precision, TDOP. In general, VDOP and GDOP are most likely to degrade the accuracy of GPS position measurements. VDOP accounts for loss of accuracy in the vertical direction caused by the angles at which the satellites being used for the position measurement are seen in the sky. If the satellites are all close to the horizon, the angles between the satellites and the receiver are all similar and VDOP can be large. In the worst possible case, if all the satellites were at the horizon, it would be impossible to make an accurate measurement in the vertical direction. A change in range to at least one satellite must occur when the receiver is moved, otherwise the receiver cannot detect that change. If all the satellites are at the horizon, no range change occurs for vertical movement of the receiver and consequently vertical accuracy is very poor. Similarly, if all the satellites were clustered directly overhead, HDOP would be large.

VDOP is important in aircraft position measurements, where height above the ground is a critical factor, especially when landing. C/A code receivers suffer from significant VDOP and cannot provide sufficient vertical accuracy for automated landing of aircraft. C/A code GPS receivers cannot guarantee sufficient vertical accuracy unless operated in a differential GPS mode.

The GPS satellites are configured in orbit to minimize the probability that a DOP can become large, by arranging the orbits to provide clusters of four satellites with suitable spacings in the sky. However, if the receiver's view of the sky is restricted, for example, by buildings, the geometry for the position calculation may not be ideal and GDOP can become large. This causes all the other DOP values to increase. Aircraft, and ships at sea always have a clear view of the sky, but automobiles often do not. C/A code receivers may revert to two-dimensional measurements (x and y) using three satellites when the sky is obstructed.

12.11 DIFFERENTIAL GPS

The accuracy of GPS measurements can be increased considerably by using *differential GPS* (DGPS) techniques. There are several forms of DGPS, all of which are intended to increase the accuracy of a basic GPS position measurement, and to remove the effects of selective availability. A second, fixed, GPS receiver at a *reference station* is always required in a differential GPS system. In the simplest forms of DGPS, a second GPS receiver at a known position continuously calculates its position using the GPS C/A code. The calculated (x, y, z) location is compared to the known location of the station and the differences in x, y, and z are sent by a radio telemetry link to the first GPS receiver. The accuracy of the C/A code position measurement can be increased from 100 m to about 10 m, with SA in effect, but this technique works well only if the two stations are close together and use the same four satellites for the position calculation.

In a more sophisticated form of differential GPS, the monitoring station at a known location measures the error in pseudorange to each satellite that is visible at its location, and telemeters the error values to users in that area. This allows other GPS users to select which satellites they want to observe, and extends the area over which the DGPS system can operate. The accuracy of a C/A code measurement can be increased to 5 m for receivers within 10 km of the reference station and to 10 m for receivers within 500 km of the reference station.

The most accurate forms of differential GPS use the relative phase of the many signals in the GPS transmissions to increase the accuracy of the timing measurements. Suppose that you could count the number of cycles of the 1575 MHz L1 carrier wave between a satellite and a GPS receiver, and that the GPS satellites are stationary for the length of time it takes to make the count at two separate locations. The wavelength of the L1 carrier is 0.19043 m, so movement of the receiver by 0.01 m directly away from the satellite would change the phase angle of the received wave by 18.9°. If the total number of cycles between the satellite and the receiver is known, and fractional cycles are measured with a phase resolution of 20°, the true distance to the satellite can be found to 0.01 m accuracy. In principle, measurements which compare the phase angle of the received L1 carriers from several GPS satellites could therefore be used to detect receiver movements at the centimeter level. This is called *differential phase* or *kinematic* DGPS.

The obvious difficulty is that we cannot count the number of cycles of the L1 carrier between the satellite and the receiver. However, we can make phase measurements and time of arrival comparisons for various GPS signals at two different locations and resolve motion between the two locations. If one of the receivers is a fixed reference station, it is then possible to locate the second GPS receiver very accurately with respect to that fixed location.

This technique is valuable in land surveying, for example, where a reference station can be set up at a known location, such as the corner of a plot of land, and the position of the plot boundary relative to that point can be measured. The same technique can be used to find the position of an aircraft relative to an airport runway so that a precision approach path can be established.

The difficulty with DGPS phase comparison measurements is that the L1 carrier has cycles which repeat every 0.19043 m, and one cycle is identical to the next. This creates *range ambiguity* which must be resolved by reference to the wavelengths of other signals. The 10.23 MHz P code transmission of the L1 carrier has a P code chip length in space of 29.326 m, which is 154 cycles of the L1 carrier. The ambiguity of the carrier waveform can be resolved within the 29.326 m length of a P code chip by comparison of the time of arrival of a particular cycle of the L1 carrier with the time since the start of the P code chip. Similar ambiguity resolution for the 29-m P code chips is possible using the length of the C/A code chip and the C/A code sequence. The length of a C/A code chip at 1.023 MHz is 293.255 m, and the length of a C/A code sequence is 293.255 km. When ambiguity resolution is applied using all of these waveforms, very small movements of the receiver can be detected and ambiguity out to 293 km can be removed. Aircraft flight paths have been tracked to an accuracy of 2 cm over distances of tens of kilometers using phase comparison DGPS techniques.

This explanation of kinematic differential GPS is oversimplified, because the satellites are moving and measurements over a considerable time are required to resolve ambiguity to the centimeter level. The P code can be used for real-time differential measurements without knowledge of P code itself, because only a comparison of the time of arrival of the code bits is required. Selective availability is not applied to the P code, so differential measurements made with the P code cannot be affected by SA.

In the *Wide Area Augmentation System* (WAAS) developed by the FAA for aircraft flying in North America, 24 WAAS receive stations continuously monitor their position as calculated from the C/A codes of all visible satellites in the GPS system. The stations also use the P code transmissions to make accurate differential measurements of the pseudorange to each visible satellite. The actual position of the WAAS stations is known very accurately from prior survey data, so each WAAS station can calculate the error in the pseudorange to each visible satellite. The 24 WAAS stations send their data to a central station with an uplink to a GEO satellite. The central station validates the data, combines all the information, and sends a sequence of pseudorange correction data to all GPS users via the satellite. The central station also determines whether any of the data is in error, and sends a warning signal called an *integrity message* to instruct aircraft not to use the GPS system, or a particular satellite, because the data are not reliable. This is an essential part of the FAA strategy for using GPS as the primary means of aircraft navigation. If the aircraft is relying on GPS information alone to determine its position, that information must have a very high reliability.

The WAAS GEO satellite transmits signals which are in a similar format to the L1 signals transmitted by a GPS satellite. A conventional GPS receiver with suitable software can extract the pseudorange error values from the WAAS satellite transmission and obtain markedly improved accuracy in its position determination. Thus no hardware changes are needed to convert a GPS receiver to use WAAS data. The GEO satellite can also be used to augment GPS satellites for position measurements, since it radiates the same signal format. The calculation of pseudorange error from the P code sequence, rather than (x, y, z) position data error from the C/A code, significantly increases the accuracy of the WAAS DGPS system.

Eventually, it seems probable that local area augmentation systems (LAAS) using differential GPS will be established at many airports to replace or augment existing ILS precision approach systems. Advanced LAAS DGPS systems have been demonstrated to achieve better than 1-m accuracy in three dimensions, with update rates sufficiently fast to control a passenger aircraft. This is sufficient to allow DGPS position data to be coupled to the aircraft autopilot so that blind landings can be made automatically in zero visibility conditions. Several demonstrations of autoland using DGPS were made in the late 1990s using Boeing 737 and 757 aircraft.

Aircraft used by overnight delivery companies will likely be fitted with GPS blind landing systems first, since cargo aircraft are subject to fewer restrictions than passenger aircraft and overnight delivery is subject to delays when airports are closed by low visibility weather. Typically, a good autoland system fitted to a large aircraft can achieve more consistent landings than a skilled pilot, so autoland may eventually become as common for landings as autopilot use is for en route operation. Weather may eventually be less of a factor in causing delays to passenger aircraft arrivals and departures.

12.5 SATELLITE SIGNAL ACQUISITION

The GPS receiver must find the starting time of the unique C/A code for each of four satellites. This is done by correlating the received signal with stored C/A codes, as in any direct sequence spread spectrum system. (See Chapter 6 for details of this process.) Usually, the receiver will automatically select the four strongest signals and correlate to those. If the geometry of the strongest satellites is poor, that is, the satellites are close together and have pseudoranges that are nearly equal, the receiver may also use several weaker signals. If the receiver is making a *cold start*, with no information about the current position of GPS satellites, or its own location, it must search all 37 possible C/A codes until it can correlate with one. Once correlation is obtained, the data stream (called the *navigation message*) from that satellite can be read by the receiver. The data stream contains information about the adjacent satellites, so once one signal is correlated, the receiver no longer needs to search through all the other 36 possible codes to find the next satellite; it can go directly to the correct code. Searching all 36 C/A codes of 1023 bits for correlation can be a slow process. In the worst case, 36 codes might have to be searched before a correlation could be obtained. However, available satellites in 2000 all had numbers between 13 and 45[5], so, on average, 16 codes might have to be searched before correlation is successful.

A direct sequence spread spectrum receiver locks to a given code by matching the locally generated code to the code received from the wanted satellite. Since the start time of the code transmitted by the satellite is not known when the receiver commences the locking process, an arbitrary start point must be selected. The locally generated code is compared to the received code, bit by bit, through all 1023 bits of the sequence, until either lock is found, or the receiver concludes that this is not the correct code for the satellite signal it is receiving.

If the starting time for the locally generated code was not selected correctly, correlation will not be obtained immediately. (This will occur with a probability of 99.9% when the timing of the locally generated sequence is selected at random.) The locally generated code is then moved forward one bit in time, and correlation is attempted again. The process is continued 1023 times until all possible starting times for the locally generated code have been tried. If the satellite with that particular C/A code is not visible, no correlation will occur and lock will not be achieved. It takes a minimum of 1 s to search all 1023 bit positions of a 1023 bit C/A code, so in a typical case, it will take at least 15 s to acquire the first satellite. Many receivers search for a given C/A code several times before moving to the next code, so several minutes may elapse before the correct C/A code is found, given no other information. Once one C/A code is found, the remaining satellites can then be acquired in a few seconds because their IDs are known from the data transmitted in the navigation message of each satellite.

Although it takes only 20 s on average to lock to the C/A code of one satellite, the receiver must find the Doppler frequency offset for at least one satellite before correlation can occur. The receiver bandwidth is matched to the bandwidth of the C/A code. The theoretical noise bandwidth of the C/A code receiver is 1.023 MHz and the velocity of the satellites is 3.865 km/s. The angle between the spacecraft velocity vector and a receiver on earth is 76.1° when a GPS satellite is at the horizon, so the maximum velocity component toward a receiver is $v_r = 928$ m/s, giving a maximum Doppler shift in the L1 signal of $v_r/\lambda = 4.872$ kHz, ignoring the effect of earth rotation. Allowing the satellite to reach an elevation angle of 5° before it is used for a position measurement limits the Doppler shift that must be accommodated by the receiver to ±4 kHz. From a cold start, the receiver must try eight Doppler frequency shifts of up to ±4 kHz in 1-kHz steps when searching for the signal from a satellite. This can increase the acquisition time of the first satellite to several minutes. Figure 12.7 illustrates the search process. There are eight possible Doppler shifts for each signal, and 1023 possible code positions, giving 8184 possible signal states that must be searched.

Once any of the GPS satellites has been acquired, the navigation message provides sufficient information about the adjacent satellites for the remaining visible satellites to be acquired quickly. The receiver may need to search in Doppler shift because the position of

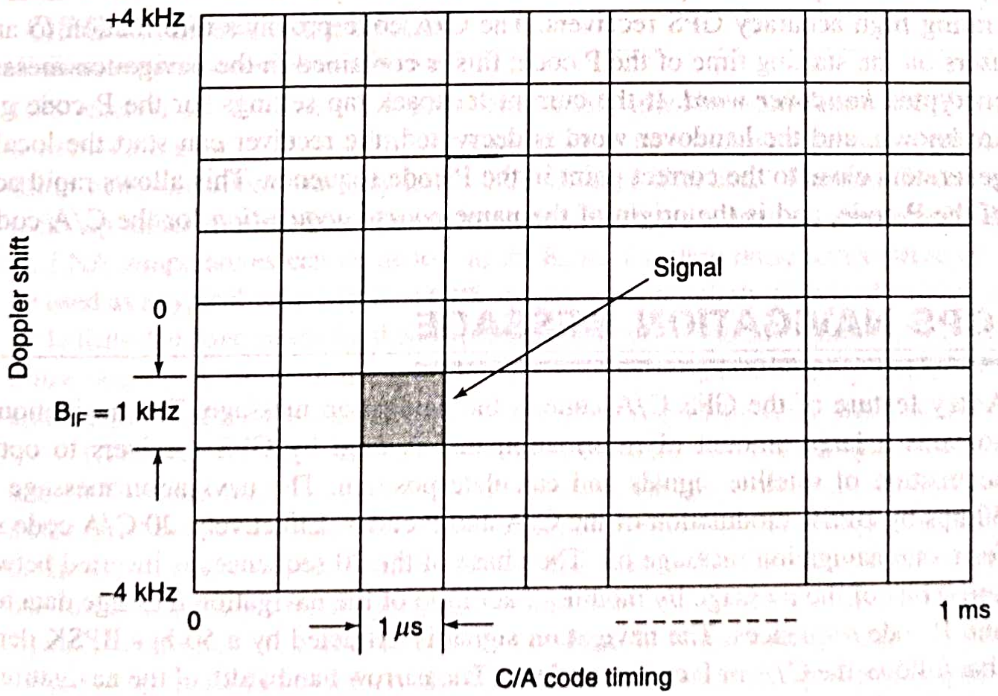

FIGURE 12.7 Code synchronization and Doppler tracking matrix.

the receiver relative to the satellites is not known, but their C/A codes are. The GPS receiver retains the information from the navigation message when switched off, and it also runs its internal clock. When next switched on, the receiver will assume that its position is close to its last known position when it was switched off, calculate which satellites should be visible, and search for those first. This greatly speeds up the acquisition process. If the receiver has been moved a large distance while turned off, a cold start may be needed.

The correlation process described above assumes that each satellite is acquired sequentially. Some lower cost GPS receivers use sequential acquisition of the satellites, and also make timing measurements sequentially, one satellite at a time. More sophisticated receivers have parallel correlators which can search for and acquire satellites in parallel. Twelve parallel correlators guarantee that all visible GPS satellites will be acquired, and start-up time is much shorter than with sequential acquisition. Accuracy is also better with parallel processing of the signals.

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

The P code for the *i*th satellite is generated in a similar way to the C/A code. The algorithm is

$$P_i(t) = XI(t) + X2(t + iT_c) \tag{12.5}$$

where T_c is the period of the *X1* sequence, which contains 15,345,000 bits and repeats every 1.5 s. The *X2* sequence is 37 bits longer. The P code repeats after 266.4 days, but is changed every 7 days for security reasons. The long length of the P code sequence makes the distance measurements unambiguous. P code sequences cannot be acquired easily because they do not repeat, a deliberate feature to prevent unauthorized users from operating high accuracy GPS receivers. The C/A code provides information to authorized users on the starting time of the P code; this is contained in the navigation message as an encrypted *handover word*. If the current feedback tap settings for the P code generators are known, and the handover word is decrypted, the receiver can start the local X Code generators close to the correct point in the P code sequence. This allows rapid acquisition of the P code, and is the origin of the name *coarse acquisition* for the C/A code.

12.6 GPS NAVIGATION MESSAGE

A key feature of the GPS C/A code is the navigation message. The navigation message contains a large amount of information that is used by GPS receivers to optimize the acquisition of satellite signals and calculate position. The navigation message is sent at 50 bps by BPSK modulation of the C/A and P codes. Effectively, 20 C/A code sequences form one navigation message bit. The phase of the 20 sequences is inverted between the 1 and 0 bits of the message by modulo-2 addition of the navigation message data to the C/A and P code sequences. The navigation signal is extracted by a 50-bps BPSK demodulator that follows the C/A or P code correlator. The narrow bandwidth of the navigation message ensures a high S/N ratio at the demodulator input and correspondingly low probability of

TABLE 12.1 GPS Navigation Message: Subframe Details

Header	Telemetry message: health of satellite, handover word
Subframe 1	Satellite clock correction data. Age of transmitted data
Subframe 2 and 3	Ephemeris for this satellite
Subframe 4	Almanac data for satellites 25 and higher. Ionospheric model data
Subframe 5	Almanac data for satellites 1–24. Health data for satellite 1–24

Note: Subframes 1, 2, and 3 repeat all data every 30 s. Subframes 4 and 5 repeat every 30 s, but transmission of the full data set requires 25 subframes over a period of 12.5 min.

bit errors in the navigation message. Satellites with elevation angles above 10° will typically give a S/N ratio of greater than 17 dB at the output of the correlator.

The complete navigation message is 1500 bits, sent as a 30-s frame with 5 subframes. However, some information is contained in a sequence of frames, and the complete data set requires 12.5 min for transmission. The most important elements of the message are repeated in every frame. The subframes contain the satellite's clock time data, orbital ephemeris for the satellite and its neighbors, and various correction factors. Details of the subframes are given in Table 12.1.

The calculation of position in a GPS receiver requires very accurate knowledge of the location of the satellite at the time that the measurements of pseudoranges are made. If the pseudorange is measured to an accuracy of 2.4 m, we must know the satellite position to an even greater accuracy, and that requires very accurate calculation of the GPS satellite orbits. By comparison, the orbit of a communication satellite does not need to be known to the same level of accuracy. As described in Chapter 2, the GPS system uses modified WGS-84 data to define the earth's radius, Kepler's constant, and the earth's rotational rate. Data on the speed of EM waves is taken from the International Astronomical Union. The WGS-84 data set also includes a very detailed description of the earth's gravitational field, which is essential for precise location of the satellites in their orbits. All of these parameters and corrections are stored in every GPS receiver, and used in calculating position.

12.7 GPS SIGNAL LEVELS

GPS receiver antennas have low gain because they must be omnidirectional. We will assume a worst-case gain of $G = 0$ dB, corresponding to an isotropic antenna. In practice, $G > 0$ dB in many directions, but may fall to 0 dB in some directions. The omnidirectional antenna picks up radiated noise from the environment, making the antenna temperature close to 273 K. LNA temperatures can be as low as 25 K, so a system noise temperature of 273 K will be used as a typical value. Typical GPS antennas are circularly polarized patches or quadrafilar helices that have carefully shaped patterns that cut off quickly below 10° elevation to minimize noise pick up from the ground. The LNA is mounted directly below or behind the antenna to avoid the increase in noise temperature caused by lossy antenna cables.

GPS satellites have an array of helical antennas that provide gain toward the earth, and 10 W transmitters, leading to EIRP values in the range 19 to 27 dBW. The C/A code transmitted by the satellite is a direct sequence spread spectrum signal, so the C/N ratio in the C/A code's RF bandwidth will be less than 0 dB. This is typical of systems that use direct sequence spread spectrum signals. The low C/N ratio of the spread spectrum signal is converted to a usable S/N by correlation of the code sequences, which adds a despreading (processing) gain to the C/N ratio. The theoretical processing gain of a direct sequence spread spectrum signal is equal to the ratio of the chip rate to the bit rate in the spreading sequence, but losses in the correlation process always make practical gains a little lower. For the C/A code transmitted at 1.023 Mbps and a 1-ms correlation time, the theoretical processing gain is 1023, or 30.1 dB. The corresponding processing gain for the P code is 40.1 dB.

The GPS receiver can pick up signals from up to 10 satellites at the same time. The RF energy from the satellite spread spectrum transmissions adds to the noise in the receiver as an interference term, I. For simplicity, in the following analysis we will assume that there are 10 GPS satellites visible, that there are 9 interfering satellites generating random signals (noise) out of which the receiver must extract the 10th signal, and that all the received signals are of equal strength. The signals from interfering satellites are treated as random noise because the Gold codes that they transmit have very low cross-correlation with the code from the wanted satellite. Noise has zero cross-correlation with the wanted signal, and the Gold codes used by GPS satellites are chosen because they closely approximate noise.

Nine interfering GPS satellites represents a worst case; in practice the number of visible satellites varies between four and ten, and the signal strengths also vary depending on the elevation angle of the satellite and the antenna pattern at the receiver. The worst case is actually when a weak signal from a satellite at a low elevation angle must be extracted from stronger signals from satellites at higher elevation angles. GPS receivers automatically select the strongest signals for processing so that the worst case can be avoided, but if the sky is partially blocked by obstructions, a weak signal may have to be used.

Table 12.2 shows the downlink signal power budget for the L1 and L2 carriers. A receiving antenna gain of 0 dB is assumed.

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

$$I = 9 \times 10^{-16}\ \text{W} \Rightarrow -150.5\ \text{dBW}$$

The thermal noise power, N, in a noise bandwidth of 2 MHz for a noise temperature of 273 K is kTB_n watts, where

$$N = -141.2\ \text{dBW} \Rightarrow 7.59 \times 10^{-15}\ \text{W}$$

The noise and interference powers must be added in watts, not in decibels:

$$N + I = 8.49 \times 10^{-15}\ \text{W} \Rightarrow -140.7\ \text{dBW}$$

Hence the worst case C/N for one C/A code signal in this scenario is

$$C/(N + I) = -160.0 - (-140.7) = -19.3\ \text{dB} \tag{12.6}$$

Similar analysis yields the $C/(N + I)$ values for the two P code signals, as shown in Table 12.3. Note that thermal noise is the major factor in setting $C/(N + I)$, since in

TABLE 12.2 Link Budget for L1 and L2 Carriers

	Carrier L1 (1575 MHz)		Carrier L2 (1227 MHz)
Code	C/A code	P code	P code
EIRP (dBW)	26.8	23.8	19.7
Path loss (dB)	−186.8	−186.8	−185.7
Receive antenna gain (dB)	0	0	0
P_r (dBW)	−160.0	−163.0	−166.0

TABLE 12.3 C/N, Noise, and Interference Budget for L1 and L2 Carriers

	Carrier L1 (1575 MHz)		Carrier L2 (1227 MHz)	
Code	C/A code	P code	P code	
T_s (dBK)	24.4	24.4	24.4	
B_n (dBHz)	63.0	73.0	73.0	
N (dBW)	−141.2	−131.2	−131.2	(thermal)
I (dBW)	−150.5	−153.0	−156.0	(nine satellites)
$N + I$ (dBW)	−140.7	−131.1	−131.2	
P_r (dB)	−160.0	−160.3	−166.0	
$C/(N + I)$ (dB)	−19.3	−31.9	−34.8	
G_{proc} (dB)	30.1	40.1	40.1	
S/N (dB)	10.7	8.2	5.3	

the worst case of interference caused by nine visible satellites, all received at maximum power, the interference power level is 9.3 dB below the thermal noise power. The C/N ratio at the receiver is 0.7 dB lower when the interference from the nine visible satellites in included. A more realistic scenario would have four satellites at the maximum receive power level and the remainder at a lower level, since GPS satellites orbit in constellations of four, with one constellation always visible, to improve the accuracy of position location measurements. Thus we would expect less than 0.7 dB degradation in the C/N ratio due to interference by other satellites' CDMA signals for almost all of the time.

The S/N at the correlator output is 10.7 dB for the C/A code and 8.2 dB for the L1 P code, using the C/N values in Table 12.3 and the theoretical processing gains for each code with no losses in the correlation and filtering of the signals. Historically, the earlier generations of GPS satellites have had transmitter EIRPs up to 3 dB higher than indicated by Table 12.3. Receiving antennas with gain greater than 0 dB also help to increase the C/N ratio, so C/A code S/N ratio can be up to 6 dB higher than the specification value of 10.7 dB for the C/A code.

The navigation message has a 50-bps bit rate, and each bit extends over 20 C/A code correlation periods. The C/A code correlator output is passed through a 50 Hz bandwidth filter which integrates the 20 pulses from the correlator to give a single message bit, in the form of a 50 bps BPSK signal. The S/N ratio of the BPSK message signal will theoretically be 13 dB higher than the S/N at the correlator output, at 23.3 dB. However, the correlation and filtering processes are not perfect and an implementation margin of several dB must be allowed. Nevertheless, the S/N ratio of the BPSK signal will be above 20 dB in most cases, guaranteeing error free detection of the navigation message.

12.9 GPS RECEIVER OPERATION

A C/A code GPS receiver must be able to correlate signals from at least four satellites, calculate time delays, read the navigation message, calculate the orbits of the GPS satellites, and calculate position from pseudoranges. The key to accurate position determination is accuracy in the timing of the arrival of the Gold code sequences from each satellite in view. All GPS receivers use a microprocessor to make the required calculations and to control the display of data. There are many different ways that this can be done, depending on the application for which the receiver is intended. The tasks of the microprocessor will not be considered here—it is assumed that once accurate timing data is available and the navigation message read that the microprocessor can complete its required tasks.

Most C/A code GPS receivers use an IC chip set that contains 12 parallel correlators. This allows the receiver to process signals from up to 12 satellites at the same time, which helps keep all the signals synchronized. Some simpler receivers use a single correlator and process four satellite signals sequentially, with consequent lower accuracy. The received GPS signals are converted to a suitable IF frequency in the front end of the receiver, and then processed to recover the C/A codes. In more recent GPS receivers, much or all of the signal processing is done digitally using DSP techniques.

The explanation of the signal processing techniques used in GPS receivers that follows is based on block diagrams that can be implemented in analog or digital form. The blocks presented in this discussion are those that would be found in an analog receiver. Most GPS receivers implement them using digital signal processing techniques (DSP). We will start the analysis by considering the signal received from the satellite at the output of the IF stage of the receiver.

The IF signal in the GPS receiver will consist of the sum of a number (up to 12) of signals from visible GPS satellites. The IF carrier signal has several BPSK modulations applied to it by the satellite, and when received on earth has been Doppler shifted by satellite and earth motion. The IF signal from N GPS satellites in view is

$$s(t) = \sum_{i=1}^{N} \{A_i C_i(t) D_i(t) \sin[(\omega_i + \omega_d)t - \phi_i(l_i) + \phi_i]\} \tag{12.10}$$

where A_i is the amplitude of the received signal.

$C_i(t)$ is the Gold code modulation

$D_i(t)$ is the navigation message modulation

ω_i is the IF frequency of the received carrier

ω_d is the Doppler shift of the received signal

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

ϕ_i is the phase angle of the transmitted signal

The key to successful measurements in a GPS C/A code receiver is to generate a signal in the receiver that is identical to the signal received from satellite i, but without the navigation data that is modulated onto the transmitted signal. When the correct signal is generated in the receiver it has the correct C/A code for satellite i, the code has the correct starting delay, and the correct Doppler shift has been applied. The locally generated signal is multiplied by the received signal, which contains several other signals from visible GPS satellites, and the output is integrated over the C/A code length of 1 ms. The result is a constant output for a period of 20 ms, corresponding to the duration of a navigation data bit. The precise matching of the locally generated signals to the received signals from four visible GPS satellites ensures that the local receiver's chip clocks and C/A code generators are exactly in sync with the received signals. When this condition is achieved, the start time of each C/A code sequence and the corresponding chip clock transition provide the high accuracy time marker that makes GPS time delay measurement possible.

The receiver must measure $\phi_i(l_i)$ in Eq. (12.10) as a time delay in order to obtain the pseudorange for each of the N satellites in view, and it must recover the $C_i(t)$ modulation by correlation. The $D_i(t)$ modulation contains the navigation message as a 50 bps BPSK modulation of the $C_i(t)$ signal. Both the $C(t)$ and $D(t)$ signals are modulated onto the carrier of the satellite signal by binary phase shift keying (see Chapter 5) and therefore have values ± 1. Demodulation of BPSK signals requires a locally generated carrier which is locked to the phase of the received carrier, and recovery of the data signal requires a bit clock that is locked to the bit rate of the received signal.

The wanted signal is buried below the receiver noise and CDMA interference. We must multiply the signal and noise by the wanted C/A code sequence to despread the signal and to bring it above the noise. The nominal bandwidth of the signal is 1 kHz after the correlator, since the 1023 bit sequence of the C/A code repeats every millisecond. However, the IF carrier can be shifted in frequency by up to 4 kHz because of Doppler effects. The receiver must therefore first search in Doppler frequency space—eight 1 kHz frequency offset steps—until a signal is found. This is done as part of the signal acquisition process by incrementing the frequency of the locally generated carrier in 1 kHz steps.

Part of a typical receiver structure for the GPS C/A code is shown in Figure 12.8. The function of the non-coherent delay lock loop is to set the frequency of the voltage controled oscillator (VCO) in the receiver to match C/A code rate of the received signal, and to align the received chip transitions correctly. GPS satellites generate all their signals from a master clock, which means that there is phase coherence between the chips, the codes and the RF frequencies of all GPS signals from a particular satellite. The delay lock loop shown in Figure 12.8 takes advantage of the coherent nature of the GPS C/A signals, so that the VCO becomes both a time reference for the C/A code signals and also the chip clock. The PN code generator in Figure 12.8 must be set to the correct code, and its start time must also be set correctly, for the loop to lock. When the IF C/A code in the receiver is correctly generated and has the correct frequency and timing, it will exactly match the received C/A code at the input to the delay lock loop.

The delay lock loop has three paths: punctual, early (half chip ahead), late (half chip behind). The delay lock loop steers the chip clock so that the punctual output can be used to drive the C/A code generator. The C/A code chip rate is generated by the VCO. The incremental process of trial and error which eventually finds the correct sequence and timing was described above. The early-late channels in the delay lock loop generate output

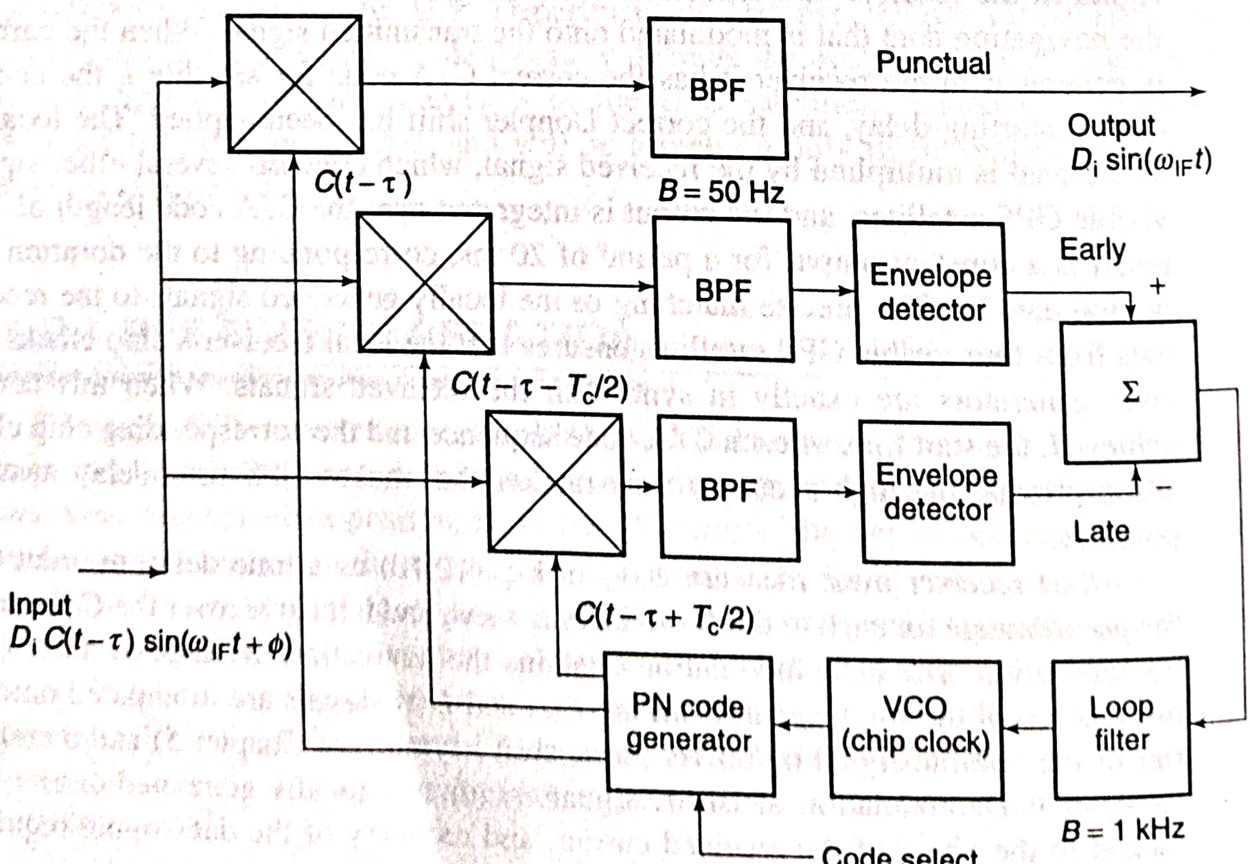

FIGURE 12.8 Noncoherent code lock loop and navigation message recovery. VCO, voltage controlled oscillator.

signals which steer the phase of the VCO so that the navigation message is recovered correctly.

The locally generated carrier that is used to demodulate the $C(t)$ signal must be Doppler shifted to match the Doppler offset of the received signal, and modulated with the correct C/A code sequence, starting at the correct time. The correct Doppler shift, code sequence, and start time are all unknown when the receiver is first switched on. The signal is buried below the noise, so it is not possible to determine the correct parameters by direct analysis of the received signal. The receiver must therefore be designed to search all possible Doppler shifts, code sequences, and code start times until an output is obtained from the correlator indicating that a satellite signal has been found. Once one GPS satellite signal has been found, information contained in the navigation message can be used to steer the receiver to the parameters needed to acquire the other visible satellites. If the receiver is turned off and then turned on again, the microprocessor memory has the last known satellite configuration stored, and can derive expected signal parameters by allowing for the time for which the receiver was off.

The output of the C/A code correlator with Doppler corrected IF frequency for the satellite signal with code number M is

$$x(t) = A_{\mathrm{m}}R(\tau_{\mathrm{m}} - \tau)D_{\mathrm{m}}(t)\sin[\omega_{\mathrm{m}}(t) - \phi_{\mathrm{m}}(l_{\mathrm{m}}) + \phi_{\mathrm{m}}] + n(t) \tag{12.11}$$

where $R(\tau_{\mathrm{m}} - \tau)$ is the autocorrelation function of the wanted code number M, and $n(t)$ is the output from cross-correlation with all other codes.

The time shift $(t_{\mathrm{m}} - \tau)$ to the correlation peak is the wanted measurement that provides the pseudorange to the satellite. The output of the correlator is a despread signal at baseband, which is modulated with the 50 bps navigation message. With the C/A code removed by the correlation process, it is a straightforward process to demodulate the navigation message D. Passing this signal through a narrow bandwidth bandpass filter improves the S/N ratio and ensures that the message is recovered without errors. The IF carrier is recovered with a special type of phased locked loop (PLL) called a Costas loop. A Costas loop compensates for the arbitrary phase of the received signal.

The despread IF carrier is BPSK modulated by the navigation message $D_{\mathrm{m}}(t)$

$$y(t) = A_{\mathrm{m}}R(\tau_{\mathrm{m}} - \tau)D_{\mathrm{m}}(t)\sin[\omega_{\mathrm{m}}(t) - \phi_{\mathrm{m}}(l_{\mathrm{m}}) + \phi_{\mathrm{m}}] + n(t) \tag{12.12}$$

The IF carrier signal is limited to remove any amplitude variations, which sets $A_{\mathrm{m}} = 1$. Then

$$y'(t) = R(\tau_{\mathrm{m}} - \tau)D_{\mathrm{m}}(t)\sin[\omega_{\mathrm{m}}(t) - \phi'_{\mathrm{m}}] + n(t) \tag{12.13}$$

The navigation message $D(t)$ is recovered by multiplying the IF signal $y'(t)$ by $\sin[\omega_{\mathrm{m}}(t) - \phi' m]$ and low pass filtering to obtain the 50 bps signal. The reference carrier for the BPSK demodulator can be derived from the output of the Costas loop. The demodulated message signal is $z(t)$ where

$$z(t) = R(\tau_{\mathrm{m}} - \tau)D_{\mathrm{m}}(t) + n'(t) \tag{12.14}$$

Provided that the correlation peak of $z(t)$ crosses the threshold and $n'(t)$ doesn't, we can recover the data message $D_{\mathrm{m}}(t)$ correctly. If everything works correctly in the receiver, the S/N of the signal $y'(t)$ is at least 17 dB, so there will be no bit errors. Even if a bit error occurs in the navigation message, it is removed when the next message is received about 30 s later.

Figure 12.9 shows a Costas loop which is often used as the demodulator for low speed BPSK signals such as the 50 bps GPS navigation message. The loop has an I channel and a Q channel driven by a VCO. The VCO frequency is set by the sum of the outputs

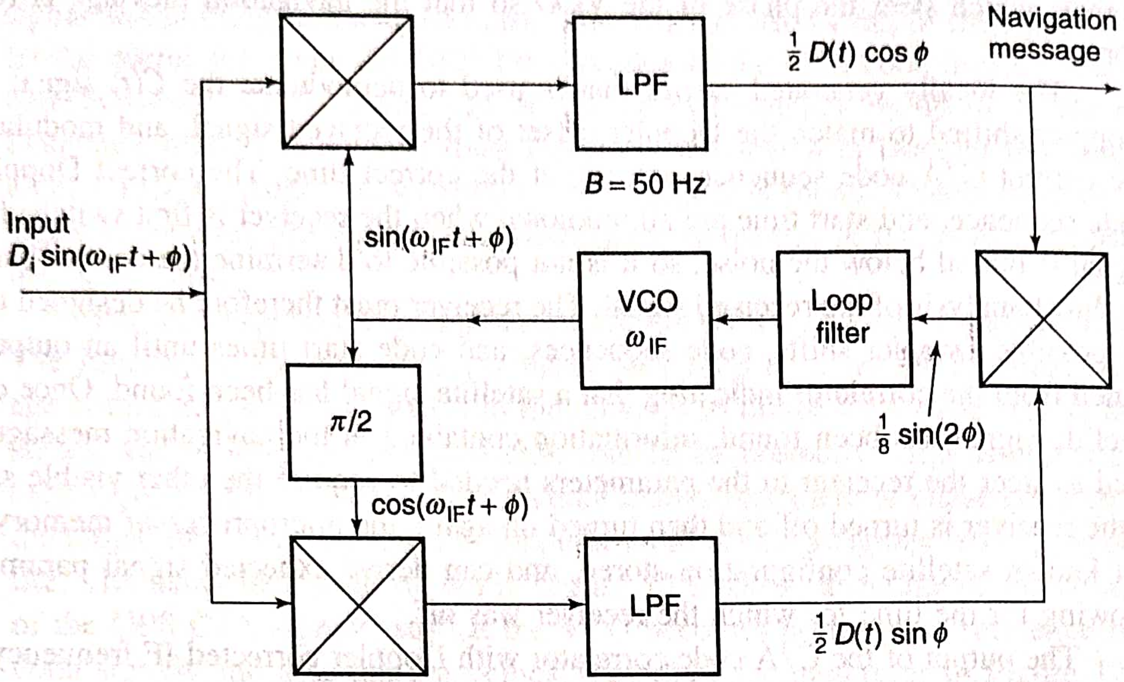

FIGURE 12.9 Costas loop. LPF, low pass filter.

from the I and Q channel detectors, which steers the VCO phase such that the I channel is in phase with the signal. The I channel output is then (ideally) a zero ISI waveform which can be integrated and sampled to recover the navigation message bits.