# Chapter 3. The Global Positioning System

3.1 The GPS Constellation

The NAVSTAR Global Positioning System (GPS) is a constellation of 24 satellites that provides users with continuous, worldwide positioning capability using the data transmitted in the GPS navigation message. The receiver uses delay lock loops (DLLs) to correlate incoming bit sequences with identical sequences generated by the receiver (see Fig. 3.1). The bit sequences are Gold codes formed as a product of pseudorandom noise sequences [Spilker, 1978]. By aligning the received and locally generated codes, the receiver determines the pseudorange to each satellite [Getting, 1993]. When the measurement is made, the receiver clock time is compared to the satellite time of transmission to determine the pseudorange. The satellite time of transmission is encoded onto the bit sequence using the navigation data.

There are six orbital planes, spaced $60°$ apart and each containing four satellites. The inclination of the orbital plane with respect to the equator is $55°$ in order to provide global coverage. At an altitude of 20,200 km, an equatorial orbit would not provide coverage above $72°$ latitude [Spilker, 1996]. Earlier Block I satellites were inclined at $63°$, but this was lowered to $55°$ for the current Block II satellites. This was done primarily because the space shuttle was planned as the launch vehicle for GPS satellites [Forssell, 1991]. However, the Challenger accident prompted a switch to Delta-II rockets for satellite launches although the orbit inclination remains at $55°$. Each orbit is nearly circular with an eccentricity close to zero (< 0.02). The orbital period of 11 hr 58 min allows for the use of Doppler frequency shift which occurs due to
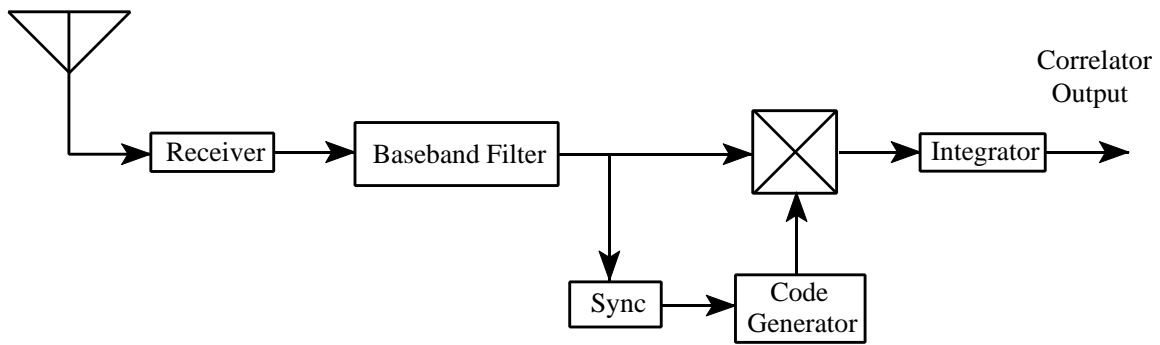
Figure 3.1 Block Diagram of a Simplified GPS Receiving System

relative motion — thus, geosynchronous orbits would not be as useful in this regard. Such

carrier phase measurements can be used along with the transmitted pseudorandom noise codes to

improve accuracy. One application of interferometry with GPS is centimeter-level surveying

[Remondi, 1984].

The design lifetime of the satellites is 7.5 years [Hurn, 1989], though some may last

longer than that. In fact, a Block I satellite launched in 1983 (PRN 12) was still operating under

healthy status in the summer of 1995. The more recent Block II satellites carry four atomic

clocks on board — two cesium and two rubidium oscillators. These are needed so that a stable

common time reference is used by all the satellites. A network of ground stations is used to

collect information that is sent to the master control station for calculation of satellite

ephemerides and modeling of the satellite clocks [Hofmann-Wellenhof, 1994]. The satellite

derives its power from solar panels whose area totals 7.25 m^2.

3.2 The GPS Signal

3.2.1 Background

The signal format for GPS is designed to satisfy a set of desired attributes which were

outlined at the conception of the system [Spilker, 1978]. Among these requirements were the

need for accurate time of arrival (TOA) measurements of the incoming signals, as well as

accurate Doppler measurements. An ionospheric group delay correction was desired. Rapid

acquisition capability was a fundamental driver in the design of the signal structure. An efficient

data channel providing the navigation message to multiple users was needed while minimizing the amount of transmitted data, allowing for small, low gain antennas to be used [Hurn, 1989]. Another key issue was susceptibility to interference (intentional or otherwise) and multipath errors.

To satisfy these requirements, spread spectrum or code division multiple access (CDMA) technology is used. In CDMA, all satellites transmit signals at the same frequency. Each satellite has its own code by which it modulates a bit stream, and this is used by a receiver to choose the satellite of interest out of background noise and other satellites transmitting in the same band. To provide the ionospheric correction, two L-band frequencies are used. The fundamental frequency (f_0) on board each satellite is 10.23 MHz, and each Block II satellite carries two rubidium and two cesium clocks. The L_1 frequency is 1575.42 MHz ($154f_0$) and the L_2 frequency is 1227.6 MHz ($120f_0$). The GPS signals are transmitted using right hand circular polarization (RHCP).

During the development of GPS, there was concern that the high accuracy delivered by this system posed a potential national security threat. Therefore, full PPS accuracy is reserved for military users. A method of limiting accuracy to civilian users (Selective Availability) was developed and implemented by the Department of Defense. The accuracy available from GPS depends on accurate time of transmission determination as well as accurate knowledge of satellite position at the time of transmission. To degrade the GPS signal, the DoD decided to dither the satellite clock and thereby introduce errors into the time of transmission determination.

In addition, the satellite orbital parameters or ephemerides are erroneous so that accurate determination of satellite position is not possible in real-time. The accurate orbital parameters are available on a bulletin board two days after the fact.

Military users are given a crypto-key which allows them to strip SA errors. Note that the P-Code is not classified; however, the crypto-key required to strip Selective Availability is classified. Civilian users do not have this crypto-key, though it is possible to track the P-code on both link frequencies. Further denial of access is accomplished by encrypting the P-Code, although some civilian receivers are still able to track these signals using advanced processing techniques.

Two positioning services are provided by GPS — the standard positioning service (SPS) and the precise positioning service (PPS). The SPS utilizes a pseudorandom noise code (PRN) of length 1023 bits known as the Coarse/Acquisition (C/A) Code. The period is 1 ms, which results in a bit rate of 1.023 Mbps ($f_0/10$). As mentioned previously, each satellite is assigned a unique code. SPS horizontal accuracy is 100 m (2d rms) with 50 m typical while the vertical accuracy is 156 m (2 rms) [Federal Radionavigation Plan, 1994]. The C/A-code is transmitted on L_1 only. The PPS utilizes a PRN code whose period is almost nine months at 10.23 Mbps. Each satellite uses a one-week segment of the P-Code. In practice, the P-code is reset each Saturday at midnight (GMT). The P-code signal is transmitted on both link frequencies. PPS horizontal accuracy is 22 m (2d rms) with 10 m typical while the vertical accuracy is 27.7 m (2 rms). The main difference between SPS and PPS is the ability to remove the intentional degradation known

as Selective Availability (SA), and the fact that the P-code bit rate is ten times faster than the

C/A-code bit rate.

The two link frequencies used by GPS allow for an ionospheric correction to be made. Note that the ionospheric correction parameters broadcast by the satellites are used in a model by single frequency users. The group delay caused by the ionosphere depends on the total electron content (TEC) in a 1 m^2 (cross-sectional area) column along the propagation path, which can be very difficult to predict. While single frequency users must apply the ionospheric model, dual frequency users can measure the ionospheric group delay for a signal from a particular satellite. Following Forssell (1991), the group delay can be approximated as:

$$\Delta t \ = \ \frac{k}{f^2} \tag{3.1}$$

where: Δt is the ionospheric group delay
          f is the frequency of the transmitted signal
          k is a constant

If two frequencies are used, then the two group delays can be differenced:

$$
\begin{aligned}
\Delta t_2 \ - \ \Delta t_1 \ &= \ k\left(\frac{1}{f_{L2}^{\ 2}} \ - \ \frac{1}{f_{L1}^{\ 2}}\right) \\[2mm]
&= \ \frac{k}{f_{L1}^{\ 2}}\left(\frac{f_{L1}^{\ 2} \ - \ f_{L2}^{\ 2}}{f_{L2}^{\ 2}}\right) \\[2mm]
&= \ \Delta t_1\left(\frac{f_{L1}^{\ 2}}{f_{L2}^{\ 2}} \ - \ 1\right)
\end{aligned}
\tag{3.2}
$$

where: f_{L1} = 1575.42 MHz

f_{L2} = 1227.6 MHz

Δt_1 is the group delay on L_1

Δt_2 is the group delay on L_2

Thus, a measure of the differential group delay (Δt_2 - Δt_1) formed by differencing L_2 and L_1 pseudoranges yields an estimate of the L_1 group delay (Δt_1).

Using CDMA allows GPS to transmit at very low power levels (see Table 3.1). In fact, the received RF signal is below the level of thermal noise in the receiver. The navigation data can be recovered because of the high coding gain associated with CDMA. Normally only the P-Code is transmitted on L_2; however, a modification can be made so that the C/A-Code is transmitted instead. The maximum received signal level does not exceed -153 dBW for a 0 dBIC, right hand circularly polarized antenna receiving the L_1 C/A Code. The signal is strongest at an elevation of $40°$ due to the shape of the satellite transmit antenna pattern [Spilker, 1996]. For the C/A Code on L_1, the power level is about 2 dB higher (-158 dBW) than the nominal minimum (-160 dBW) when the satellite is at $40°$ elevation using a 3 dB linearly polarized receiving antenna [ICD-GPS-200, 1991].

Table 3.1   Minimum Received Signal Power for GPS (0 dBIC, Right Hand Circularly Polarized Receiving Antenna)

GPS Signal Strength	C/A-Code	P-Code
L_1 (1575.42 MHz)	-160 dBW	-163 dBW
L_2 (1227.6 MHz)	-166 dBW[*]	-166 dBW

[*]Not Implemented During Normal Operation

3.2.2 The GPS Signal Structure

The L_1 signal has two orthogonal components, given by [Spilker, 1978]:

$$S_i(t)_{L1} = A_p P_i(t) D_i(t) \cos(\omega_1 t + \phi) + A_c G_i(t) D_i(t) \sin(\omega_1 t + \phi) \qquad (3.3)$$

where:
$\omega_1 = 2\pi(1575.42 \text{ MHz})$
$P_i(t)$ is the P-code at 10.23 Mbps
$G_i(t)$ is the C/A-Code at 1.023 Mbps
$D_i(t)$ contains the navigation message at 50 bps (Amplitude ±1)
A_c, A_p govern the relative amplitude of the in-phase and quadrature
components
ϕ represents phase noise

The navigation message is 1500 bits long (30 s) and contains five six-second subframes [ICD-GPS-200] as shown in Table 3.2. A telemetry word (TLM) and handover word (HOW) begin each subframe. It takes twenty-five frames or 12.5 minutes to obtain almanac data for each satellite vehicle (SV).

The C/A-Code is stronger than the P-Code ($A_c>A_p$) by 3-6 dB typically. It is a Gold code generated as a product of two pseudorandom noise sequences, one of which is offset by a certain number of bits in order to achieve unique codes for each satellite [Spilker, 1978]. At 1.023 Mbps the 1023 bit C/A-Code has a period of 1 ms. The P-Code is also generated as a product of two pseudorandom noise sequences, one of which is offset by 0 to 36 bits. Thus, 37 unique P-Code sequences are available. At 10.23 Mbps, the period would be nearly nine months long if the code were not reset weekly. Once the C/A-Code is acquired, the handover word (HOW) is used to acquire the P-Code. At present, civilian receivers can track the encrypted P-code with a low

35

Table 3.2 Organization of the Five Subframes of GPS Navigation Data

	Word Number (Each Word is 30 bits)		
	1	2	3-10 (240 bits)
1	TLM	HOW	Satellite Clock Correction Parameters
2	TLM	HOW	Satellite Ephemeris Data
3	TLM	HOW	
4	TLM	HOW	Almanac and Health (SV #25-32), Special Messages
5	TLM	HOW	Almanac and Health (SV #1-24), Almanac Ref. Times

signal to noise ratio using averaging techniques and the fact that the encryption code is 20 times

slower than the P-code bit rate. The Selective Availability correction is not available except to

authorized users.

The L_2 signal contains the P-Code only:

$$S_i(t)_{L2} \quad B_p P_i(t) D_i(t) \cos(\omega_2 t \quad \phi) \tag{3.4}$$

which contains terms similar to Eq. 3.3 except that here $\omega_2 = 2\pi(1227.6 \text{ MHz})$.

3.3 Satellite Geometry

The GPS satellite geometry relates position errors to range measurement errors. Consider

equation A.10-A.11 with m pseudorange measurements. Then $\partial \underline{Y}$ is an m x 1 vector, $\mathbf{H}$ is an

m x 4 matrix, and as usual $\partial \boldsymbol{\beta}$ is a 4 x 1 vector. If we think of $\partial \boldsymbol{\beta}$ as a zero-mean vector

containing the errors in the estimated user state, then we are interested in the statistics of $\partial \boldsymbol{\beta}$

because that will characterize the expected position errors. Using the generalized inverse of $\mathbf{H}$,

we find the covariance of $\partial \boldsymbol{\beta}$ [Brown & Hwang, 1992]:

$$
\begin{aligned}
cov(\partial \boldsymbol{\beta}) \;=\; E[\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T] \;&=\; E[(H^T H)^{-1} H^T \partial \underline{Y} \partial \underline{Y}^T H (H^T H)^{-T}] \\
&=\; (H^T H)^{-1} H^T E[\partial \underline{Y} \partial \underline{Y}^T] H (H^T H)^{-T} \\
&=\; (H^T H)^{-1} H^T cov(\partial \underline{Y}) H (H^T H)^{-T}
\end{aligned}
\tag{3.5}
$$

Now we have the covariance of $\partial \underline{Y}$, the pseudorange errors. These are assumed to be

uncorrelated, Gaussian random variables. As such, they are statistically independent which

results in a diagonal covariance matrix. Furthermore, the range measurement errors are assumed

to have the same variance (σ_r^2) for each satellite. So now we have:

$$cov(\partial \underline{Y}) = \sigma_r^2 I \qquad (3.6)$$

which results in:

$$E[\partial \underline{\beta} \partial \underline{\beta}^T] = \sigma_r^2 (H^TH)^{-1} H^TH(H^TH)^{-T}$$

$$= \sigma_r^2 (H^TH)^{-T} \qquad (3.7)$$

It can easily be shown that H^{T}H is symmetric, so the transpose is unnecessary:

$$cov(\partial \underline{\beta}) = \sigma_r^2 (H^TH)^{-1} \qquad (3.8)$$

Now the relationship between range measurement errors and position errors becomes clear. Let

$\mathbf{G} = (H^TH)^{-1}$ so that $cov(\partial \underline{\beta}) = \sigma_r^2 \mathbf{G}$. Using $B = ct_B$ and expanding Eq. 3.8 we get:

$$
\begin{bmatrix}
\sigma_x^2 & cov(X,Y) & cov(X,Z) & cov(X,B) \\
cov(Y,X) & \sigma_y^2 & cov(Y,Z) & cov(Y,B) \\
cov(Z,X) & cov(Z,Y) & \sigma_z^2 & cov(Z,B) \\
cov(B,X) & cov(B,Y) & cov(B,Z) & \sigma_B^2
\end{bmatrix}
=
\begin{bmatrix}
G_{xx} & G_{xy} & G_{xz} & G_{xB} \\
G_{yx} & G_{yy} & G_{yz} & G_{yB} \\
G_{zx} & G_{zy} & G_{zz} & G_{zB} \\
G_{Bx} & G_{By} & G_{Bz} & G_{BB}
\end{bmatrix}
\sigma_r^2 \qquad (3.9)
$$

The elements of $\mathbf{G}$ give a measure of the satellite geometry called the dilution of precision (DOP)

[Milliken and Zoller, 1980]. Various DOP values can be calculated from the diagonal elements

of **G**. For example:

$$\sigma_x^2 + \sigma_y^2 + \sigma_z^2 + \sigma_B^2 = (G_{xx} + G_{yy} + G_{zz} + G_{BB})\sigma_r^2$$

$$\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2 + \sigma_B^2} = GDOP\,\sigma_r$$

(3.10)

where GDOP is the Geometric Dilution of Precision defined as:

$$GDOP = \sqrt{G_{xx} + G_{yy} + G_{zz} + G_{BB}}$$

(3.11)

A high GDOP is associated with poor satellite geometry, as seen in Fig. 3.2. Other DOPs are as

follows:

$$PDOP = \frac{\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}}{\sigma_r} = \sqrt{G_{xx} + G_{yy} + G_{zz}}$$

$$HDOP = \frac{\sqrt{\sigma_x^2 + \sigma_y^2}}{\sigma_r} = \sqrt{G_{xx} + G_{yy}}$$

(3.12)

$$VDOP = \frac{\sigma_z}{\sigma_r} = \sqrt{G_{zz}}$$

$$TDOP = \frac{\sigma_B}{\sigma_r} = \sqrt{G_{BB}}$$

where: PDOP is the Position Dilution of Precision
HDOP is the Horizontal Dilution of Precision
VDOP is the Vertical Dilution of Precision
TDOP is the Time Dilution of Precision

The most common measures of geometry are HDOP and VDOP. Typical values using all satellites in view would be 1.0 for HDOP and 1.5 for VDOP. These values can be much higher if the geometry is poor, but VDOP is almost always higher than HDOP. This will be discussed in more detail in Chapter 4.

**Good Geometry
(Small PDOP)**

Earth

**Poor Geometry
(High PDOP)**

Earth
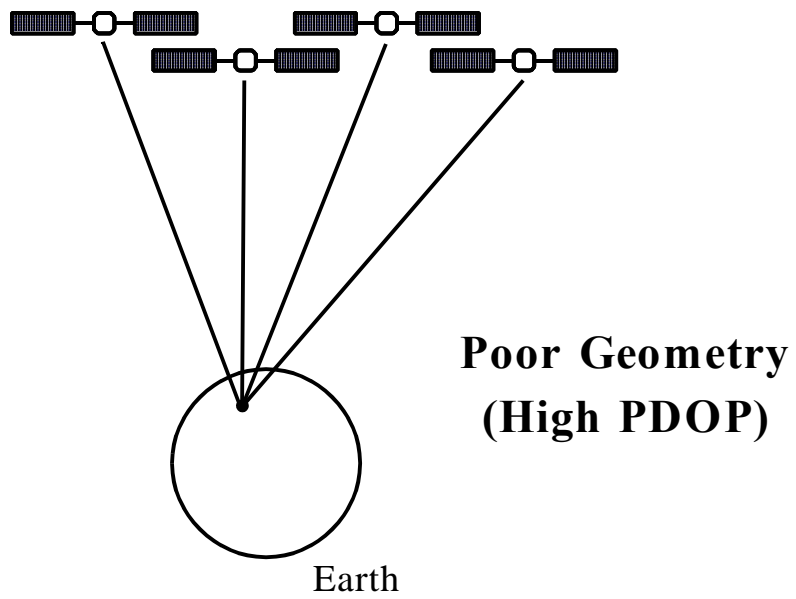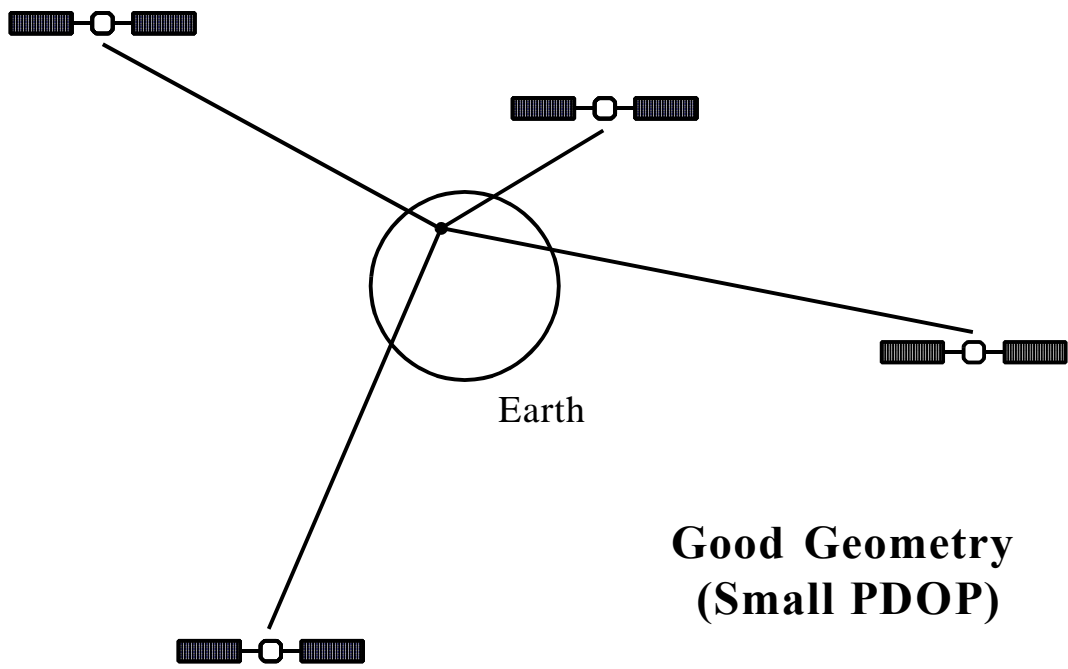
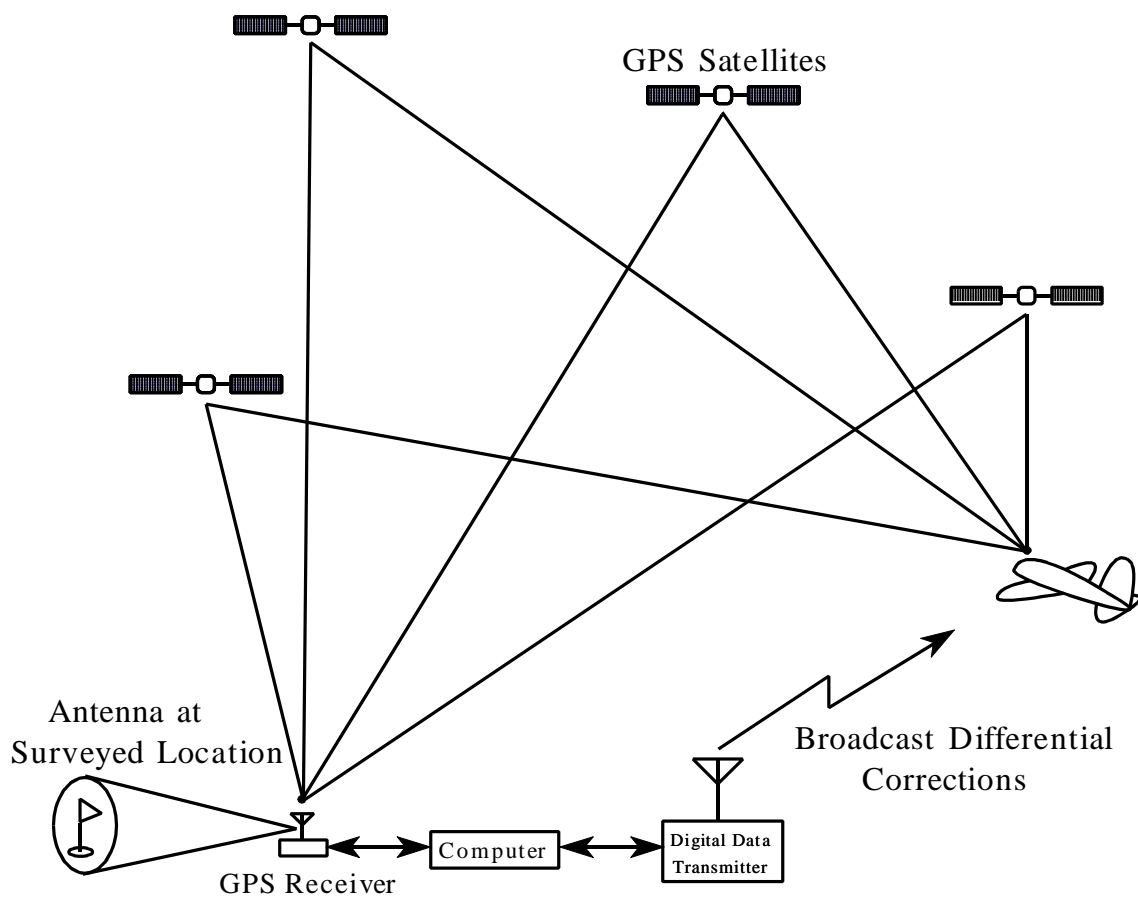**Figure 3.2 Illustration of Satellite Geometry**

3.4 Differential GPS

3.4.1 Basic Differential GPS Operation

Appendix A shows how a stand-alone user can determine position using GPS pseudoranges. As mentioned before, SPS accuracy is 100 m horizontal (95%) and 156 m vertical (95%) [Federal Radionavigation Plan, 1994]. To achieve better accuracy in the presence of Selective Availability and other errors, differential techniques are used to eliminate errors that are common to two users. The basic concept is illustrated in Fig. 3.3.

In Differential GPS (DGPS), a reference station is typically located at a known, or surveyed position. By comparing measurement data to the truth, the ground station can generate corrections for users in a local area. This can be done in a post-processing environment, or using a data link if high accuracy is needed in real-time. In this way, horizontal position can be improved to 5 m or better [Kremer, Kalafus, Loomis and Reynolds, 1990], [Waid, 1991], [Braasch,1992]. This represents better accuracy than stand-alone PPS users achieve.

Differential processing removes errors that are common to both the ground station and the user. Errors that are highly correlated between the two include ionospheric group delay and selective availability. Ionospheric errors are approximately the same for users in close proximity [Diggle, 1994]. Selective Availability errors cancel between two users if the receivers make simultaneous measurements. Because SA is a slow-changing error, it should be approximately equal between the ground and air receivers even though the times of transmission are slightly different for ground and air received signals.

42

GPS Satellites

Antenna at
Surveyed Location

Broadcast Differential
Corrections

GPS Receiver

Computer

Digital Data
Transmitter

Figure 3.3 Calculating Position Using Differential GPS

One error source that is not common is tropospheric delay. However, unlike the ionosphere, the troposphere can be modeled fairly well, though not perfectly without extensive meteorological equipment on board. Using standard temperature, barometric pressure, and humidity, users at different altitudes can eliminate much of the tropospheric path delay. Error sources that are difficult to eliminate are noise and multipath. Multipath represents the dominant error source in DGPS systems [Waid, 1991]. In some applications multipath errors can be reduced by averaging, although multipath is not necessarily zero-mean [van Nee, 1991]. Here we are limited by noise and multipath on the code phase measurements, necessitating the use of carrier phase measurements for improved accuracy.

3.4.2 Differential GPS Using Carrier Phase Tracking

While five-meter accuracy is adequate for many applications, aircraft navigation during final approach and landing requires greater precision. In order to achieve sub-meter accuracy during periods of low visibility, further processing of the GPS signal is required. Differential carrier phase tracking techniques are used to improve accuracy.

Carrier phase advance can be measured very precisely as a function of Doppler shift [Wells et. al., 1986]. Carrier phase measurements are much less noisy than code phase measurements and therefore provide better accuracy. It should be noted that the use of narrow correlators can reduce the noise in the code phase tracking loops [van Dierendonck, 1992]. The difficulty in dealing with carrier phase measurements is that there exists an unknown number of

integer wavelengths between the receiver and the satellite when the signal is acquired. In order to determine the precise range to the satellite, this initial ambiguity must be resolved.

Interferometric techniques using GPS were first implemented in surveying applications [Counselman & Gourevitch, 1981]. The problem is simplified because both receivers are stationary and therefore the baseline between them is fixed. In kinematic applications, one receiver is moving and the baseline is constantly changing. This requires "on the fly" ambiguity resolution [Hwang, 1991]. In order to achieve precise location in real-time, the integer ambiguities must be resolved quickly and reliably.

In practice, ambiguity resolution is performed using carrier phase double differences (DDs). This is readily understood by examining the components of a carrier phase measurement (given in terms of wavelengths):

$$\Phi(t) = \phi_{sat}(t) - \phi_{rec}(t) + N + S_n + f \cdot t_B + f \cdot t_{sat} - \phi_{iono} + \phi_{tropo} \qquad (3.13)$$

where: $\Phi(t)$ is the range measurement in wavelengths
$\phi_{sat}(t)$ is the transmitted carrier phase in fractions of a wavelength
$\phi_{rec}(t)$ is the received carrier phase in fractions of a wavelength
N is the unknown integer number of wavelengths
S_n is measurement noise
f is the carrier frequency (1575.42 MHz for L_1)
t_B is the receiver clock bias
t_{sat} is the satellite clock bias
ϕ_{iono} is the phase advance due to propagation through the ionosphere
ϕ_{tropo} is the phase delay due to propagation through the troposphere

It is interesting to note that the carrier phase is *advanced* as the signal travels through the ionosphere. This accounts for the minus sign in front of the ϕ_{iono} term. The opposite takes place in measuring the code phase; here the ionosphere causes a group delay.

The single difference can be formed either by differencing two measurements made by the receiver (between-satellite single difference) or by differencing the measurement to the same satellite by two receivers (between-receiver single difference) as shown in Fig. 3.4 [Wells et. al., 1986]. In the latter case we have:

$$SD = \Delta\phi_{rec} + \Delta N + \Delta S_n + f \cdot \Delta t_{bias} \qquad (3.14)$$

where SD is the single difference in terms of carrier wavelengths. Notice that ϕ_{sat} and t_{sat} have dropped out. Here it is assumed that the receivers are in close proximity so that ϕ_{iono} drops out as well. The residual tropospheric errors are assumed to be modeled and are also ignored. Thus the number of unknowns has been reduced.

With two receivers and two satellites, the carrier phase double difference can be formed. Following [Diggle, 1994], two single differences are formed in accordance with Eq. 3.14:

$$SD^1 = \Delta\phi_{rec}^1 + \Delta N^1 + \Delta S_n^1 + f \cdot \Delta t_B$$
$$SD^2 = \Delta\phi_{rec}^2 + \Delta N^2 + \Delta S_n^2 + f \cdot \Delta t_B \qquad (3.15)$$

The double difference is formed by subtracting one measurement from the other (see Fig. 3.5):
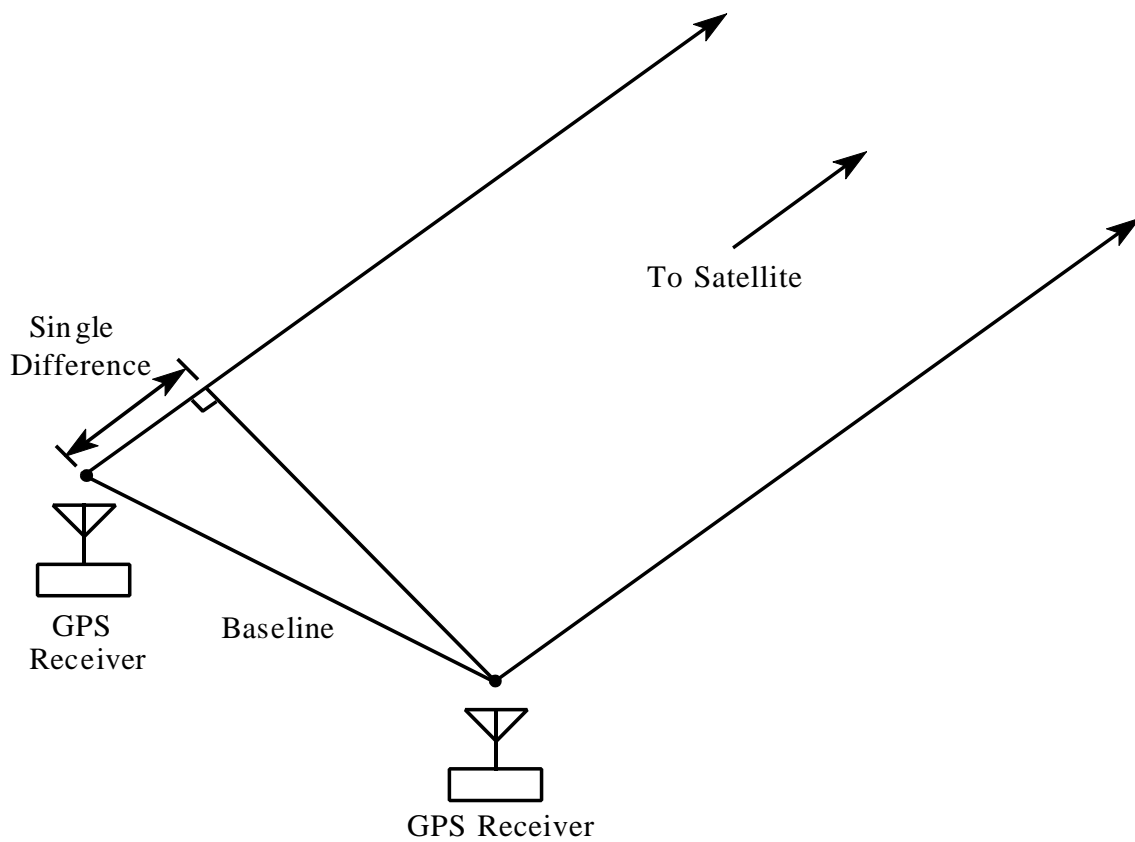
Figure 3.4 Illustration of the Single Difference

$$DD = SD^1 - SD^2$$

$$= \Delta\phi_{rec}^{12} + \Delta N^{12} + \Delta S_n^{12}$$

(3.16)

The combined receiver clock bias term drops out leaving a combined measured carrier phase term ($\Delta\phi_{rec}^{12}$), a combined integer ambiguity term (ΔN^{12}), and a combined noise term (ΔS_n^{12}).

To solve for the unknown baseline between two receivers, this integer ambiguity term must be resolved.

The ambiguity resolution process has been reported on extensively in the literature [7, 9, 10, 19, 22, 31, 44, 45, 57]. Position accuracy on the order of fractions of a wavelength can be achieved once the ambiguities are resolved. At the GPS L_1 frequency of 1575.42 MHz, this means position accuracy on the order of a few centimeters is possible. In practice the wide lane observable is often used to reduce the search for the correct set of ambiguities. In widelaning, the L_1 (1575.42 MHz) and L_2 (1227.6 MHz) signals are differenced to form a carrier phase observable with an effective wavelength of 86 cm. There is a slight sacrifice in accuracy but there is a gain in the robustness of the ambiguity search algorithm.

To Satellite 2

To Satellite 1

SD 1

SD 2

GPS Receiver

Baseline

GPS Receiver
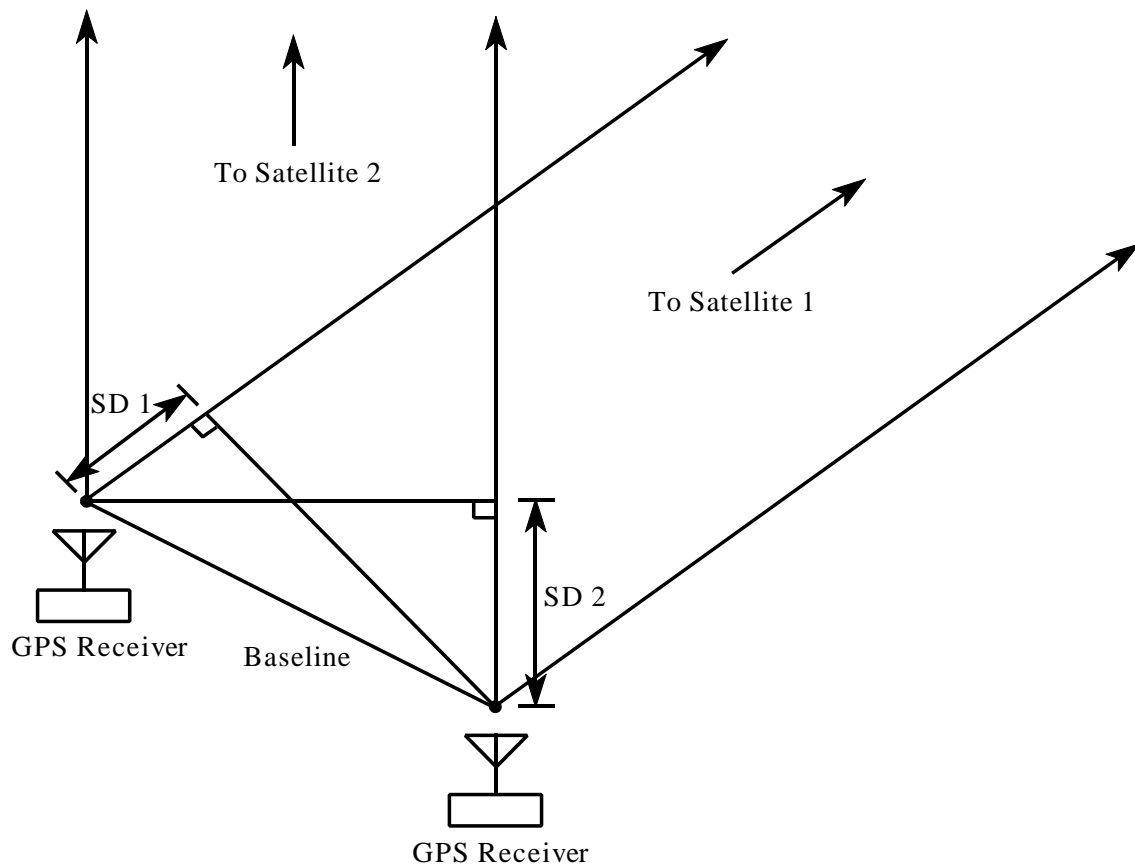
Figure 3.5 Illustration of the Double Difference