Principles of GPS

A Brief Primer on the Operation of the Global Positioning System

Carl Carter
Revision 1
February, 1997

ALLEN OSBORNE ASSOCIATES

756 LAKEFIELD ROAD
WESTLAKE VILLAGE, CA 91361-2624
TEL: (805) 495-8420
FAX: (805) 373-6067

Copyright Notice and Disclaimer

Table of Contents

Introduction

The Global Positioning Satellite (GPS) system was established by the United States Department of Defense (DoD) to provide a real-time navigation system for the US military. Since its inception it has grown to provide not only world-wide, all-weather navigation, put precise position determination capabilities to all manner of users. The resulting precision available exceeds any previously attainable without large expenditures of time and resources. This introduction will provide a brief description of how the system works and how it may be used.

The operation of the GPS system is divided into several topics. An attempt has been made to cover each topic in such a way that a new user will have sufficient information to understand the critical elements of the system, and so that an engineer will have an introduction to the major topics sufficient to suggest where additional study may be required. Section I, The Elements of the GPS System, describes the logical breakdown of the system. Section II, Principles of Operation, gives an overview of the general theory of the system's operation. Section III, The GPS Signal, and Section IV, The Navigation Message, give greater details to the implementation of elements described in Section II. Section V, GPS Errors, describes the types and sources of errors that affect the operation of a GPS system, and explains terms commonly found in GPS literature. Section VI, Differential GPS, and Section VII, Applications of GPS, explain some of the ways GPS can be used.

I. The Elements of the GPS System

The GPS system consists of three "segments" called the Control Segment, the Space Segment, and the User Segment. Proper operation of each of these three segments results in accurate, reliable operation of the entire system.

The Control Segment is composed of the main control center located at Falcon Air Force Base, near Colorado Springs, Colorado, USA, and several monitoring and control stations located around the world. These stations monitor the satellites, report the results to the main control center, and relay the control signals generated in Colorado back to the satellites. The Control Segment stations are the only ones which transmit to the satellites. The information they send to the satellites provides for positioning the satellites in orbit, provides data to be broadcast in the satellites' navigation messages, and generally provides control of the satellite operation. Part of the satellite broadcast data includes a health status. The Control Segment is responsible for detecting satellites that are not broadcasting properly, or that are not in the proper orbit, and commanding the satellites to identify themselves as unhealthy when circumstances warrant. This allows the Control Segment to keep results obtained from using the system consistently within operating specifications.

The Space Segment is composed of a constellation of satellites orbiting approximately 20,000 km (about 12,500 miles) above the Earth. The full constellation is defined as 24 satellites, but there may be more or fewer active at any one time. The satellites are arrayed in 6 separate orbits, each inclined about 55° with respect to the equator, with 4 slots per orbit designated to hold a satellite. The orbit is traversed in about 12 hours. With a full constellation, receivers located on most spots on the Earth can see at least 6, and sometimes as many as 12 of the satellites at any one time.

The User Segment is the term given to all of the receivers listening to the satellites at any time. There is no organization to the User Segment, but for any user, it consists of the receiver currently in use and its associated antenna. User receivers are passive -- they need only listen to the Space Segment and not broadcast anything, thus making the system accessible to any number of users at one time without users interfering with each other.

While all three segments operate at one time, the typical user is basically unaware of the Control Segment, and only concerns himself with the operation of his own receiver and the satellites actually visible at his location during his time of use. Further, limitations in individual receivers may make the user aware of only some of the satellites visible at his location, since the receiver may only select a few of them to monitor.

II. Principles of Operation

Broken down to the simplest terms, the satellites orbiting above the Earth simply broadcast their location and the current time. The receivers listen to several satellites (how many will be discussed below), and from the broadcasts determine what time it is and where the receivers are located. The principles, of course, require much more detail, but this the essence.

Each satellite broadcasts two signals consisting of carrier waves that undergo phase changes that occur in a defined pattern at very precise rates and at exact times (see Section III below). A receiver generates a copy of the phase-change pattern and moves it back and forth in time, attempting to correlate it with signals it receives. If the signal it is trying to correlate with is being received, at some point the received pattern and the internally generated pattern will match. The correlator circuit will then generate a large output. This pattern match and associated correlator output constitute lock-on to a satellite, and provides a pattern generator in the receiver that is working exactly in step with the received signal. Knowing how much this generator was shifted in time tells the receiver when the signal arrived at the receiver with respect to its own internal clock. If the receiver could determine how its clock was adjusted with respect to true GPS time, it would then know exactly how long it took the signal from the satellite to reach the receiver. When the receiver multiplies this time by the speed of light, it knows how far it is from the satellite.

In addition to transmitting a specific phase-change pattern that is unique for each satellite, additional data is also added to the signal. This data comprises the Navigation Message (see Section IV below). It includes the current time to the nearest second, and the information needed to compute the location of the satellite at the time of transmission. Using this information, the receiver can set its clock to the correct second, and compute the current position of the satellite. It now knows how far it is from the satellite, and where the satellite is. Using simple geometry, the receiver now knows it is somewhere on the surface of a sphere centered on that satellite with radius equal to the distance from the satellite.

Let’s look at the process in an end-to-end scenario. Several satellites are broadcasting their patterns -- each one unique -- which are arriving at the antenna of a receiver. Each pattern arrives at a different time determined by the relative distance between the receiver and the satellite sending the pattern. The receiver searches for specific satellites by generating and shifting the pattern for each satellite that might be broadcasting. Once matches are found, the receiver can compute the distance, called pseudorange, to each satellite. If the receiver’s clock is precisely coordinated to GPS time, the receiver could immediately compute its position using simple algebra.

Unfortunately, the receiver’s clock is usually not set exactly to GPS time. Thus, the pseudorange consists not only of the time it took the signal to travel to the receiver, but also an amount that represents how far the receiver clock and GPS time differ. This is called clock offset, and represents a fourth unknown (in addition to the receiver’s x, y and z position). Clock offset could be either positive or negative since the receiver clock could be either ahead of or behind GPS time. Pseudorange is measured in units of time. Because we know that the signal traveled to the receiver at the speed of light (about 300,000,000 meters per second), we can convert it to a distance simply by multiplying it by that number. Similarly, clock offset is measured in units of time and can also be converted to distance as well. This distance or time error is common to all of the pseudoranges since the receiver uses the same clock to measure all pseudoranges.

When a receiver acquires a satellite, the receiver monitors the navigation message from the satellite. Part of the data contained in the navigation message is the current GPS time, expressed in seconds. GPS time is the number of seconds since midnight between January 5 and 6, 1980. Thus, the receiver is able to set its own time indication to the exact whole second (the receiver computes fractions of a second later). Another part of the navigation message is a set of numbers called the ephemeris, that together describe the satellite's orbit in space, and where the satellite is in that orbit at a particular time. The receiver computes the exact location of the satellite in space from the ephemeris and the current time. The result is a set of **x**, **y** and **z** coordinates where the satellite was when the signal was transmitted. These values tell the position of the satellite with respect to a coordinate system defined by the World Geodetic System 1984 (referred to as WGS84). The **origin** of this coordinate system is near the Earth's mass center, and its **z** axis matches the mean spin axis of the Earth. **+z** is towards the North pole; **+x** emerges from the Earth on the Greenwich meridian at the equator (just south of Ghana, and west of Gabon, in the Atlantic Ocean). The **+y** axis emerges at the equator on the 90° East meridian (at a point in the Indian Ocean southeast of Sri Lanka and west of Sumatra), thus defining a right-hand coordinate system.

At this point the receiver has the location of each satellite, and the pseudorange to that satellite. Using appropriate math the receiver computes its position (x, y and z) and clock offset (Δt). To understand how this works, let's look at it graphically. To make it easier to visualize, we will use a two-dimensional solution. The three-dimensional solution works exactly the same, but with the added **z** factor. Refer to Figure 1. Points A, B and C are the locations of three satellites in the x, y coordinate space of the diagram. Radii d1, d2 and d3 represent the pseudoranges we measured from each satellite (here shown in the distance form). Thus, we would define our position as located on the intersection of the three circles centered on each satellite with radius equal to the respective pseudoranges. But the three circles do not meet at a point. They intersect to form a triangle with arcs for sides (in some cases, they could even miss each other entirely).

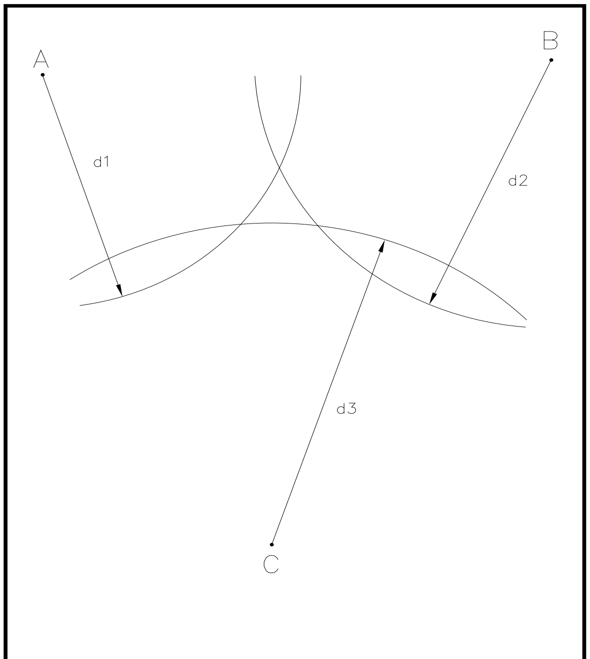

figure 1

Now refer to Figure 2. In this case we have added a small amount, t, to each pseudorange. The result is that we have adjusted each pseudorange by the same amount, **t,** causing the circles to meet at a point. The coordinates of this point represent our position, and **t** represents our clock offset. As a result of this process, we not only know our position, but we also know the correct time (fractions of a second) within the resolution of our code pattern shifter. Time resolution is typically to fractions of microseconds, resulting in a time determination that is more accurate than about any other method generally available. In fact, GPS receivers designed specifically to adjust atomic clocks yield time determinations that match UTC to within 10 nanoseconds!

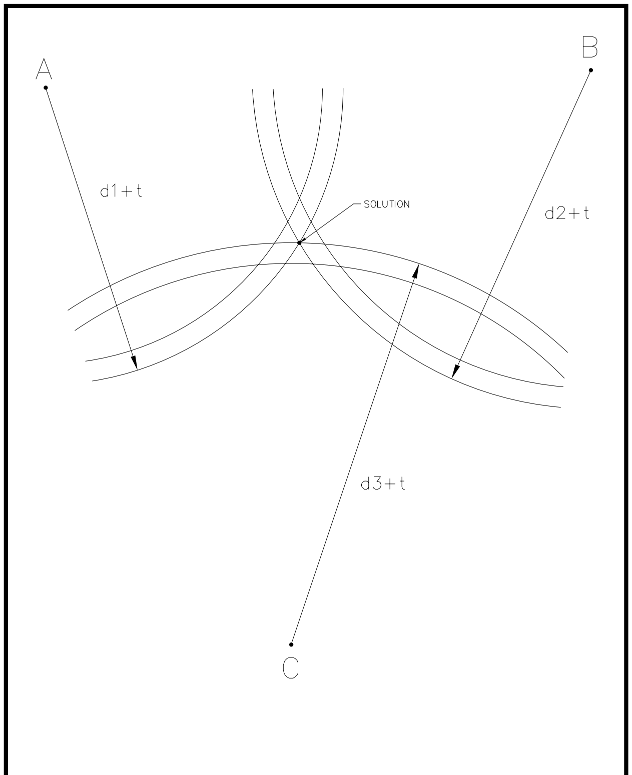

figure 2

An interesting element of the position determination process becomes apparent here. Note that if the receiver is a long distance from the antenna, the satellite signals must travel that distance inside a cable to reach the electronic circuits that measure the pseudorange. As a result, the measured pseudorange increases by the time required to travel the distance represented by the cable length. In addition, signals tend to travel slower than the speed of light inside cables (in some cables, at less than two thirds the speed of light). This factor also increases the time for the signal to reach the receiver. However, since the signals from all satellites travel this same distance, the effect is to add the same amount of time delay to all signals. Now recall that pseudorange is the sum of the time it takes the signal to travel to the receiver, and the clock offset of the receiver. Of course the distance from each satellite to the antenna is unique. The extra distance from the antenna to the receiver, and the clock offset of the receiver, are exactly the same in all measurements. When the receiver computes the solution, these two constant terms become merged into a single value referred to as clock offset. The causes the x, y and z position coordinates to be referenced to the common point of the antenna. Thus, using a longer or shorter antenna cable will not affect the position determined, but will only affect the computed clock offset. Further, there exists a single point within the antenna where the antenna detects each signal. This point is called the antenna phase center. Since it typically is displaced from the mounting point of the antenna by several centimeters, manufacturers of precision surveying receivers routinely publish the location of the phase center with respect to some convenient point accessible on the housing of the antenna. This lets surveyors measure the distance between a station being surveyed and the antenna housing, and later relate that measurement to the antenna phase center, thus adjusting the survey results for the actual distance between the station and the antenna phase center. Another point of interest specifically to those doing very high precision surveys, such as geodetic surveys, is that the phase center of an antenna is often different for different frequencies. The GPS system transmits two frequencies called L1 and L2 (described below). Thus, published phase center displacements for an antenna may give values for both L1 and L2 to allow the surveyor's post-processing software to compensate for the difference.

III. The GPS Signal

All GPS satellites broadcast on the same two frequencies. The primary signal is broadcast on what is referred to as L1 frequency, which is 1,575.42 MHz. The signals are broadcast using spread-spectrum techniques, which allow many signals to coexist on the same frequency, and for receivers to detect and separate the different signals from each other. The L1 signal is modulated with two information signals called C/A (for Coarse and Acquisition code) and P (for Precise code). In addition, the satellites also broadcast a copy of the P code on the second frequency called L2, which is 1,227.60 MHz. (Note that later satellites may add additional codes to L1 and/or L2, and may also add extra frequencies.)

Spread-spectrum modulation basically consists of the carrier signal being repeatedly inverted, that is, having its phase shifted 180°, in a specific pattern. The C/A code pattern is generated by a hardware signal generator consisting of a pair of 10-bit shift registers with feedback connections in them, whose outputs are combined by an exclusive-OR gate. The resulting digital sequence is referred to as a Pseudo-Random Number, or PRN, sequence. The generator produces a pattern that is exactly 1,023 bits long, which then repeats. By starting both the shift registers at a defined starting point, and by combining the resulting outputs with a phase shift between them (that is, the output of one register is delayed by some number of bits from the output of the other), several unique codes can be generated. The GPS system defines 36 specific phase shifts to be used, resulting in 36 unique codes (called Gold Codes) that could be transmitted by satellites. Since the satellite number is represented in the navigation message by only 5 data bits, only 32 of these 36 codes are actually used. The others are reserved for other uses, such as ground transmitters. The bit rate of the generator used to modulate the carrier is referred to as the "chipping" rate, and each bit is referred to as a "chip." For GPS satellites, the C/A code chipping rate is 1.023 MHz. Since the code is 1,023 bits long before it repeats, the code repeats every 1 msec.

To receive a spread-spectrum signal, the receiver must know the desired PRN sequence. It generates its own copy of the sequence, and applies it to the output of a down-converter and detector. The receiver then shifts the pattern in time looking for a match with what appears to be noise coming from the detector. The match is made in a circuit called a correlator, which produces an output that corresponds to the degree of match between the two signals. When the receiver's code matches the received signal, there is a large rise in the magnitude of the correlator's output. To search for a transmitter, the receiver first adjusts the internally generated pattern in time, chip-by-chip, until an indication of matchup has occurred, then shifts the pattern by fractions of a chip until the correlator output is maximized. At this point the internal pattern generator is generating a code in exact step (at least to the resolution of the pattern shifter) with the received signal.

For greater precision in determining the time it takes a signal to travel to the receiver, a second signal is generated and transmitted on the same frequency. This second carrier is 90° ahead of the carrier with the C/A code, but it is of a lower amplitude. It is modulated with a PRN sequence called the P (for Precise) code. The P code has a chipping rate of

10.23 MHz, so it is 10 times the rate, and thus, precision, of the C/A code. In addition, the P-code sequence is much longer than the C/A code - it does not repeat over a complete week. This makes it harder to acquire without the initial time setting afforded by the C/A code being acquired first (in fact, this is why the C/A code has the term “acquisition” in its name). A feature called Anti-Spoofing (A-S) can be activated by the US DoD to prevent the intentional deception of receivers by use of a phony, or “spoofing” transmitter. The result of A-S being turned on is that the P code is hidden by an encryption scheme. The P code thus encrypted is called Y code. Figure 3 shows how the C/A code and P (or Y) code modifies the carrier sine wave. While the P-code carrier also is phase shifted 180° by its bit pattern, since it is lower in amplitude than the C/A code carrier and 90° out of phase with it, the effect of combining the two carriers is for the output signal from the satellites to appear to shift by about 70° when P-code bits change.

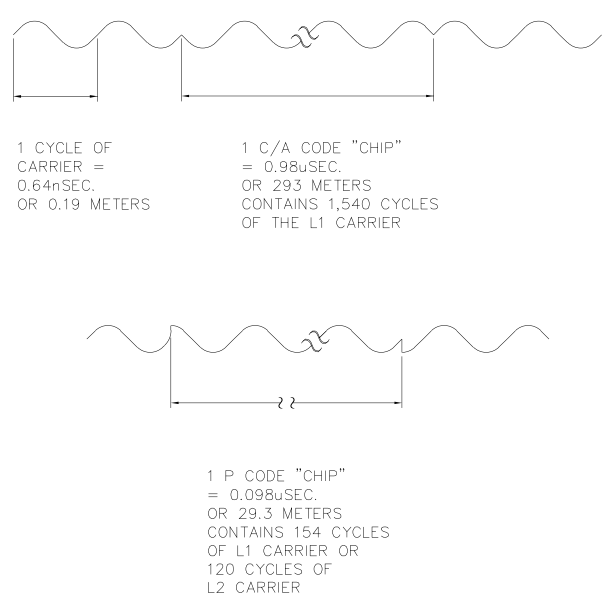

figure 3

Data being sent on the carrier is represented by either inverting or not inverting the PRN code, so that at the receiver the correlator will generate either a positive or negative correlation output. The data rate is usually much slower than the chipping rate so that it does not interfere with the integration that is done as part of the correlation process. Data on the GPS signals, called the Navigation Message, are modulated at a nominal 50 bits per second rate.

Each satellite contains multiple atomic clocks, and the carrier and modulation signals are timed precisely to the clocks. Thus, at exactly the start of a second as defined by the master GPS timing, each satellite's signal is crossing zero or passing an integral multiple of one of the phase changes, and its modulation (both C/A and P codes) are also starting bits. In fact, the L1 and L2 frequencies have been chosen so that they relate to each other and to the chipping rates in a coherent manner. The basic timing is provided by the 10.23 MHz frequency of the P code. L1 is exactly 154 times this frequency, and L2 is exactly 120 times this frequency, so a single P chip consists of 154 cycles of the L1 or 120 cycles of the L2 carrier. The C/A chip rate is composed of 10 P chips, and the actual navigation message data rate is defined as exactly 20 copies of the 1023-bit C/A pattern. Thus, a properly working satellite has all the elements of the signal locked to one reference frequency, and the phase of the carrier, the phase of the chips and the data all align with transitions occurring on 0° boundaries of the unmodulated carriers.

The coherency of the transmitted signal provides yet another method of determining pseudorange; this is referred to as carrier phase. Once a receiver has determined its own clock offset, it can determine the actual start of a second. The received signal will differ from this point in time by some integral number of carrier cycles, plus a fractional part of a cycle. Just as with the C/A code, this offset is due to the time it takes the signal to travel from the satellite to the receiver. The fractional cycle can usually be determined to about 1 part in 1000 using current technology. The integer number of cycles, however, is subject to some ambiguity, but can be determined using a process called ambiguity resolution. Since the wavelength of the carrier is about 0.19 meters, resolving this to 0.1% (one part in one thousand) yields a pseudorange measurement that has a resolution of about 0.2 mm (this is less than 8/1000”). For purposes of illustration, the P code chip length (the equivalent of wavelength) is about 29 meters, and the C/A code chip length is about 290 meters. Assuming the same 0.1% resolution on these as on the carrier waves, this implies resolution of 2.9 cm for P code (a little over 1”), and 29 cm for the C/A code (about 11”).

IV. The Navigation Message

The data modulated onto the C/A and P codes consist of several types of information. The data are packaged into 30-bit words that consist of 24 data bits and 6 parity bits. Words are grouped together into groups of 10 called a subframe. Each subframe is thus 300 bits long, of which 240 bits are data and 60 bits are parity. Subframes take 6 seconds to transmit at 50 bps. There are 5 subframes defined, numbered 1 through 5. The satellite transmits a set of all 5 subframes in 30 seconds, then begins to transmit another set. The contents of the subframes change over time as noted below. Figure 4 illustrates the navigation message and its components.

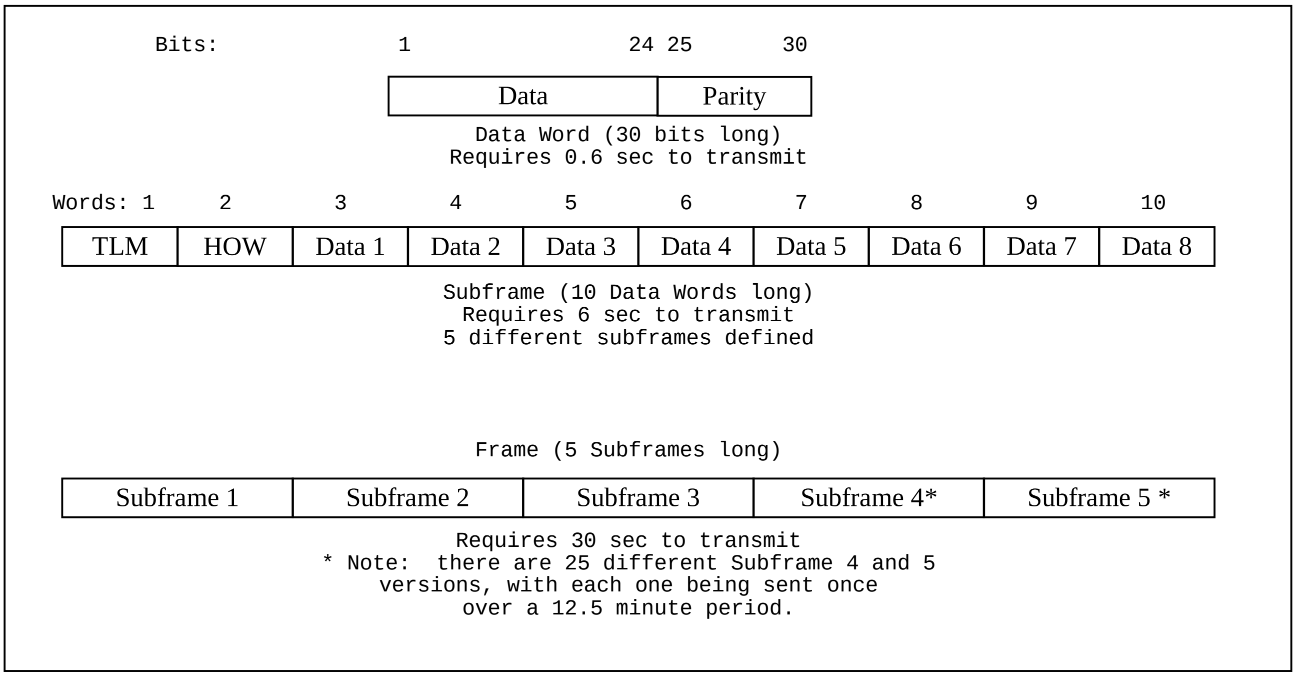

Figure 4

Each subframe starts with a pair of data elements called the Telemetry Word (TLM) and the Handover Word (HOW). The TLM provides a standard bit-pattern preamble that can be used to detect the start of a subframe, plus administrative status information such as data upload status. The HOW contains the GPS system time (referred to as Z-count) that corresponds to the start of the next subframe, and an identification of which subframe number this is. The TLM and HOW together take up the first two words of every subframe. All information discussed below about the 5 subframes relates only to the remaining 8 data words.

Subframe 1 contains information that can be used to compute a correction term for the satellite's clock. Even though the satellites have multiple atomic clocks on them, the clocks do drift. This drift is monitored by the Control Segment stations, and a second-order curve fit is made to it. The coefficients of the expression describing this curve are reported in subframe 1 so that the users can compute the current modeled clock error and thus improve their own navigation.

Subframes 2 and 3 contain the ephemeris data for the satellite. The ephemeris is presented in a format called “Keplerian elements plus secular drift terms and harmonic coefficients.” With these parameters, the user is able to accurately compute the satellite’s location at a specific time to an error of less than 0.3 meters. The data are updated about once an hour (once every 2 hours for newer satellites), with the ephemeris data known to be valid for about four hours.

Subframes 4 and 5 are used in a different manner from subframes 1-3. Among other data, they contain what is referred to as almanac data. Almanac data consists of a truncated set of ephemeris data, reduced so that it will fit in a single subframe, with corresponding reduction in accuracy. Unlike the other subframes which repeat exactly the same data for about an hour or two, these two subframes change every time they are sent. And instead of sending data only about the transmitting satellite, these subframes contain data about all satellites in the constellation. The purpose of the almanac data is to allow a receiver locked onto one satellite to find out about all other satellites in the constellation. Subframe 5 contains the almanacs for satellites 1 through 24 (satellite numbers correspond to the specific PRN Gold Code that satellite is using as its C/A code chipping signal as described in Section III), each broadcast in sequence in successive versions (called “pages”) of the subframe. After sending 24 almanacs, subframe 5 then contains a page that reports on the health of those 24 satellites. This makes 25 unique pages of subframe 5, each sent once over a 12.5 minute period, after which it repeats. Subframe 4 contains pages that contain the almanacs for satellites 25 through 32, and a health page for those satellites. In addition, there are 16 other pages defined for subframe 4 that are either used for other purposes, or reserved for later features that might be added to the GPS system. This results in 25 unique pages for subframe 4 as well. Almanac data may be used for several days, even for weeks, since the purpose is not navigation, but simply finding the satellite.

V. GPS Errors

The process of transmitting, receiving and detecting the GPS signal is a physical process which, like any other physical process, contains sources for errors. Some of the errors are obvious: the satellite clock is not exactly correct, even when the broadcast correction terms are used to adjust it. The location of the satellite in space is not necessarily correct since it is determined by observations made on the ground, and the ephemeris values only yield a solution accurate to about 30 cm. And the receiver computing its own position can only resolve the received signals to some specific precision determined by the wavelength of either the carrier (for carrier phase measurements) or the code bit length (for code pattern matching), and the resolution of the code or phase shifter in the receiver. Further limitations occur in the receiver based on the precision of the computations, where mathematical processes may truncate or round values rather than carrying them out to their last possible decimal place.

Some other error sources appear when we look at the physical process of a signal traveling through space to the receiver. For example, the signal is transmitted by a satellite traveling at a high rate of speed in space. Since it is unlikely that the receiver is traveling at the same direction and speed, there will be a Doppler shift of the signal that affects the effective wavelength of both the code and carrier waves. And the signal must travel through the ionosphere above the Earth, which has an effect of shifting the signal and bending its path. It also travels through the troposphere (the lower layer of the atmosphere where most weather occurs), which also affects the signal's path and speed. As the signal nears the receiver, some of it may reflect off of the ground, water, or buildings, water towers, signs, etc., located near the receiver, and reach the antenna after traveling a greater distance than the signal that arrives directly from the satellite (a phenomenon called multipath).

Finally, the receiver is also prone to errors that can be detected. The clock is unstable, causing individual ticks to occur not truly regularly, but with some "wobble" between them. Most receivers compute the pseudorange to a satellite several times per second, and average the measurements. If the clock in the receiver is "noisy," that is, not ticking at a uniform rate, the time during which each individual measurement is made can vary. This would result in each observation being made over a time span that is unique, and the resulting average could contain errors caused by this effect. In addition, all receivers detect noise along with the real signals, and that affects the received signals, further degrading them. Satellites closer to the horizon tend not only to be weaker, and thus more prone to noise, but their signals are more prone to multipath. It is for this reason that some receivers allow the user to set an elevation mask, that is, an angle below which satellites will not be tracked.

The result of all of these and other error sources is that the computed pseudorange is an estimate with a possible error whose magnitude can be computed using standard statistical methods. So when we compute our location with respect to a satellite, instead of finding ourselves on a sphere as described earlier (shown as the circular arcs in the two-dimensional model of figures 1 and 2), we really find ourselves in a space located between two concentric spheres -- one with radius equal to our pseudorange plus the error, one with radius equal to the pseudorange minus the error. This will be illustrated in two dimensions in the next figure.

The actual geometry of the satellites in the sky also has an effect on the accuracy of results. Returning to a 2-dimensional representation for ease of illustration, let’s see how the geometry affects our results. Figure 5 shows a case where we are only considering two satellites. Instead of a single arc, the distance from each satellite is shown as three arcs. The center, bolder arc is the same one we computed before -- representing the computed pseudorange. The two lighter arcs surrounding the center arc represent the magnitude of the error we estimate. Thus, the true range from the satellite is really somewhere in between the two outer arcs for each satellite, and therefore we are located somewhere inside the shaded area where these two spaces intersect. Now look at Figure 6. In this example the errors and ranges are exactly the same, but the two satellites are located closer together. Note how the geometry of the satellites causes the area where we might be located to grow. Obviously, if the two satellites were at the same point in the sky, the presence of the second satellite would not help our position determination at all!

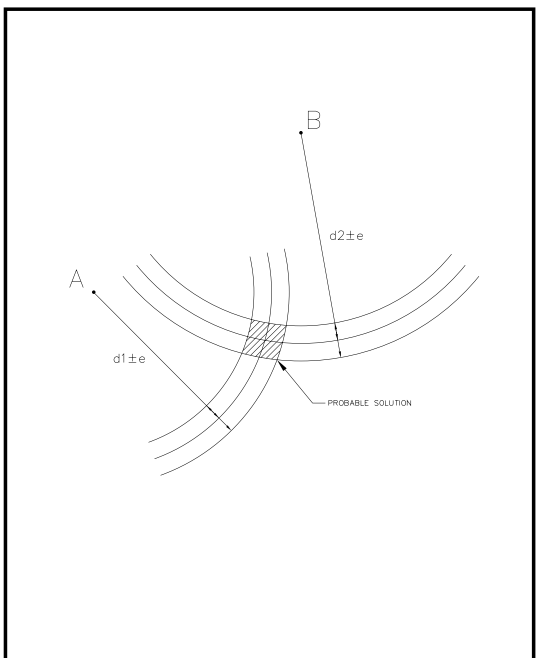

figure 5

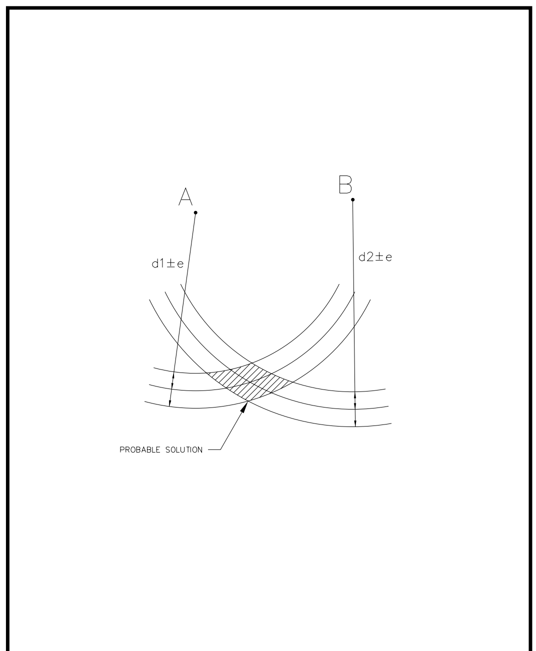

figure 6

The effect of this geometry on our overall error computation results in what is called a Dilution of Precision, or DOP. If we had such a thing as an ideal geometry, with satellites in all possible directions, we would have a DOP value of 1.0. In a more realistic situation, with 6 to 12 satellites visible, and all of them above the horizon, the DOP value rises. When true position and clock offset are computed, the error on these values and others computed from them can be determined by multiplying the composite error of the observations by the appropriate DOP.

The overall DOP term is called Geometric Dilution of Precision, or GDOP. GDOP may be broken down into two components: one related to the receiver's position (Position DOP, or PDOP) and one related to the time determination (Time DOP, or TDOP). While PDOP is related to the satellite geometry, as discussed above, TDOP is strictly dependent on the time bases in the receiver and all the satellites. Thus, it is a function only of the number of satellites being tracked. GDOP, PDOP and TDOP are related orthogonally by:

$$GDOP^2 = PDOP^2 + TDOP^2$$

The PDOP value is further found to have two components: horizontal (HDOP) and vertical (VDOP), again related to PDOP orthogonally. Here one of the limitations on the GPS system becomes apparent. The horizontal component is basically affected by how the satellites are dispersed in azimuth about the receiver. If all satellites are bunched up in a single direction, the HDOP will be larger than if the same number of satellites were evenly spaced around the horizon. But consider VDOP. It is dependent on the elevation of the satellites, as you might expect. But since satellites below the horizon cannot typically be seen by a receiver, by necessity all satellites that we use are bunched up in the space from the horizon upwards. This simple fact explains why receiver manufacturers will specify larger error values for their receivers for vertical position determinations than they will for horizontal positions. For completeness, here is how PDOP, HDOP and VDOP relate:

$$PDOP^2 = HDOP^2 + VDOP^2$$

One final error source must be noted. Since the original design of the GPS system was for the US military, there was concern that it might be used by an adversary to guide weapons. For this reason, there was a decision to put an intentional error source into the transmitted signals that would limit the availability of the system to users. This was done by such factors as altering the satellite's clock, or by altering the broadcast clock correction terms or other ephemeris terms. A user computing position using the C/A code from a fixed location would find the position solution moving over time such that the resulting position determinations would have about 100 meters 2d RMS variation. At the same time, an authorized user with access to the appropriate technology could still use the P code and navigate as accurately as before. This process is referred to as selective availability, or SA. Don't confuse it with anti-spoofing (A-S), which is the encryption of the P code into a Y code. While the SA makes real-time position determination less precise, the use of post processing with one or more reference receivers can remove most of the effects. Real-time differential GPS tracking can similarly remove these effects. Both methods are described in Section VI.

VI. Differential GPS

Earlier we presented a discussion of some of the error sources affecting GPS measurements. To summarize, there are sources associated with the satellites, sources associated with the signal traveling from the satellite to the receiver's antenna, and sources associated with the receiver. Differential GPS was developed as a means of providing real-time correction of several of these errors.

Consider the makeup of a pseudorange measurement in a receiver. The receiver has determined the difference in time between when a signal left a satellite and when it was received by the receiver. This includes the actual time of travel of the signal, plus the effects of error such as those originating in the satellite, in the path of the signal from the satellite to the receiver, and in the receiver, as discussed in Section V.

Now consider the receiver's position determination process. Recall that the receiver uses the broadcast ephemeris to determine the location of the satellite. As noted before, this value is also subject to errors; thus defining one additional error source. Therefore, even if all of the errors in measuring the pseudorange could be corrected in some manner, there would still be an error element in the resulting position.

In normal survey applications, the data from the reference receiver and from the receiver(s) that survey the points of interest, are combined in a post-processing environment, and mathematical techniques are used to remove the common errors. The result is a position that can be repeated to within as little as 1 or 2 mm. But this is only after collecting a large number of observations and performing extensive math. An alternative exists that works in real time, but with less spectacular results.

The Differential GPS (DGPS) process works using a reference receiver, as in the post-processing technique above. That receiver is placed in a location which is known very well. This might be a benchmark to be used as a starting point for a survey, the end of an airport runway or the entrance to a harbor that is critical for navigation, or some other similar location. The receiver either surveys the location, or is told the location through data entry by an operator, and it saves this position and assumes it to be true. Next, the reference receiver makes normal observations and computes pseudoranges just as any other receiver would do. It also listens to the broadcast ephemeris, and combines the current time and broadcast ephemeris to compute the satellite's location. Using the computed satellite position and its own known location, it computes the mathematical, or model, range to the satellite. If there were no errors in the broadcast ephemeris or in the satellite's location, this model range would represent what the pseudorange would be expected to be. If we compare the observed pseudorange and the model range we will nearly always find they differ, due to the errors from all the sources previously noted. The reference receiver computes this difference for each satellite, and reports all of the values as corrections over some form of data link (radio broadcast, wire connection, etc.) to any other receiver wanting to use them. The other receivers, called User receivers, or simply users, take the values and apply them to their own pseudorange observations prior to computing their own position.

The pseudoranges observed by users contain the same types of errors as those observed by the reference receiver. Of the errors, those which are introduced by the satellite are identical at both receivers. Those caused by the signal path, including ionospheric and tropospheric effects, are usually similar at both receivers, but not exactly the same. The errors caused by the receiver are of course unique to each receiver. Let's examine these errors in more detail.

Satellite errors, generally consisting mainly of clock error, are common to any observer. Clearly, the correction reported by the reference receiver contains the effects of this error. Thus, when the user receiver applies the corrections, the effects of the satellite errors are removed from the solution.

Errors in the pseudorange caused by the signal's passing through the ionosphere tend to be on the order of several meters to tens of meters. However, since the satellite is about 20,000 km above the Earth, and the ionosphere is also quite high above the Earth, the path of the signal through the ionosphere is generally quite similar for receivers which are no farther than a few hundred kilometers apart. As a result, while the errors due to signal path through the ionosphere differ somewhat, the corrections generally correct of these errors. The troposphere's effects are not as common between two receivers. First, the troposphere is much lower, extending from the surface of the Earth upwards. Thus, the signal's path through the troposphere differs much more as the distance between the two receivers increases. Second, the troposphere can vary more dramatically over short distances. The types of features in the troposphere that affect signals include the moisture content. The presence of a large cloud in one signal path, and the absence of that cloud in another could cause considerable difference in the effects. Fortunately, the total effect of the tropospheric variations tends to be on the order of parts of a meter to a few meters. Thus, while the corrections sent from a reference to a user receiver will contain effects of the troposphere, there is likely to be only partial accounting for these errors, with the benefit reducing as distance increases. Certainly, the effects may be quite different to the point that corrections of tropospheric effects are probably useless within some few tens of kilometers. The other major signal-path error, multipath, is quite unique. Since this error amounts to multiple copies of the same signal appearing at a single point in space, but with different times of arrival, the effect is one of signal interference. It is well known that such effects tend to vary greatly over a distance of less than a single wavelength of the signal, or less than 20 cm for the GPS L1 carrier! Obviously, multipath errors cannot be corrected for reliably by the broadcast corrections. This is mitigated by placing the reference receiver at a site with good satellite visibility to ensure a strong direct signal, and one that is protected from obvious multipath sources, such as on top of a building where other sources of reflected signals are below the antenna. Good antenna designs help greatly reduce the effects of multipath sources near the horizon and below, thus reducing the multipath error effects on the reference receiver's computed corrections.

Because corrections are more effective closer to the reference receiver, reference receivers are often located near harbor entrances or at major airports so that vehicles navigating to such locations have the least error in the most important portion of their navigation.

Errors originating within the two receivers used for a differential GPS data link include errors due to the clocks of the receivers, noise on the received signals, and limitations of the pattern shifting process. The corrections from the reference receiver will contain these errors. When the user receiver applies the corrections, this will have the effect of transferring reference-receiver errors to the user receiver. However, some of these errors tend to be common to all signals observed by the reference receiver (for example, clock offset errors), and thus will be included in all corrections. The user receiver will thus find the error to be the same as common terms such as those caused by using a longer or shorter antenna cable -- that is, they will become lumped into the clock offset portion of the solution rather than affecting the x, y and z position coordinates. It is for this reason that when a receiver is using differential corrections, it should not use a mixture of observations with corrections and observations for which there are no corrections present; use of such a mixture would preclude the reference receiver's common errors from being placed solely in the clock offset portion of the solution and would result in a position with greater error.

One interesting side effect of this process involves errors in the reference receiver's position. Such errors are directly propagated since they are not common to all satellites, but rather affect each satellite depending on the associated geometry. For example, if the reference receiver's position is assumed to be farther north than it really is, the effect on satellites to the north of the receiver will be in the opposite direction of that for satellites to the south; and satellites directly east or west of the location would have no different corrections due to such an error. As a result, the user receiver would find its position displace from "truth" by exactly the same amount and in the same direction as the error in the reference receiver. While this may be bad, in fact it could also be beneficial. Consider that a survey is being done relative to a benchmark whose position is published, but is in error with respect to the GPS system. If the position of this benchmark were entered for the reference station, the error in the published position would be propagated to the user receivers, and the result would be that all surveyed points would be proper with respect to the benchmark, without any conversions being necessary! Similarly, aircraft landing at an airport or ships navigating into a harbor would find themselves computing positions that agree with the local charts regardless of the presence of any errors in those charts.

Any error associated with the broadcast ephemeris will be seen at the reference receiver only in the model range, since the observations themselves are made without regard to the broadcast position of the satellite. However, their effect will be seen in the corrections, and will be accounted for in the user receiver when the position is computed since the user receiver will at that time use the same incorrect position of the satellite to convert pseudoranges into positions. Thus, all errors originating on the satellite (including those caused intentionally by selective availability, or SA) will be removed in the DGPS process.

In practice, there are errors which differential techniques cannot correct for, and even errors which are introduced by the process (such as errors in the reference receiver's observations that are not common to all satellites). But the overall effect is quite significant. Just using the pseudorange derived from the C/A code, a stationary receiver will generally compute its position as a point which moves over time by more than 100 meters. When a nearby reference station's corrections are applied, that range of movement will generally be much less than 10 meters, and even less than 2 meters when the receivers are relatively close. This real-time process makes position determinations quite accurate in most navigation modes (but still not as accurate as post-processing for surveying applications).

The above description describes what could be referred to as code differential GPS, since the observations and associated corrections are made from pseudoranges derived from the C/A or P code on the satellite carrier. Another form of differential can be developed that looks at the carrier phase. The higher resolution of the carrier wave makes the application of such corrections yield position solutions that vary by only a few centimeters. This form of differential is referred to as Real Time Kinematic, since it allows the performance of the ambiguity resolution process used for kinematic navigation to be performed in real time with immediate results.

VII. Applications of GPS

GPS receivers have been used in numerous applications. Most relate either to the ability of the receiver to determine the position of the antenna in space, or to the ability of the receiver to determine time to great precision.

The ability of GPS receivers to determine time to great precision makes them ideal clock drivers. In fact, several manufacturers have adapted receivers to provide precise time pulses out, or to produce the time pulses and then compare them to those from an external source such as a Cesium or Rubidium clock, and to provide output that tells how the two differ. These latter receivers can be connected into control systems that use the differences reported to steer the atomic clock until it comes into close agreement with the GPS time. Such receivers are used by time standards services in several countries around the world. The best receivers permit clock steering to within a few nanoseconds of GPS time. In addition, the US National Aeronautics and Space Administration (NASA) uses GPS receivers as part of their Deep Space Network so that precise time references are available to use to measure the signals from spacecraft traveling to other planets, thus permitting increased precision in determining their positions in outer space.

The ability to determine position is the most widely used feature of the GPS system. This can be used in several ways. Certainly a vehicle such as a truck, ship or aircraft can have a receiver that constantly keeps track of position. This is how many of the new navigation systems for automobiles -- those that show the vehicle’s location on a map display -- work. Some trucking companies connect receivers into radio systems that use satellite communications to allow one master control center to track all of the trucks in the fleet anywhere they travel. And emergency vehicles have been fitted with receivers so that the dispatcher can always know which vehicle is closest to an emergency. Ships at sea and aircraft flying in bad weather can use GPS anywhere on Earth to keep track of position and speed, regardless of visibility or time of day. Even satellites have been equipped with GPS receivers to help correlate scientific measurements with location of the craft.

Some aircraft have been equipped with two or more GPS antennae, with the signals from all antennae sent to one receiver. The receiver is thus able not only to compute position, but by measuring the difference between the signals from each of the antennae, to compute the attitude of the aircraft. When the aircraft has an inertial navigation system (INS - a navigation system using gyroscopes and accelerometers to sense changes in a vehicle’s attitude and velocity) connected into the navigation package, the two systems can greatly improve overall navigation reliability. GPS is not affected by drift over time, but can exhibit some variations between individual measurements due to noise. In contrast, INS is quite precise in measurements over a short range of time, but its errors accumulate over time, resulting in larger errors as the time between external updates increases. Together, the two technologies complement each other, with INS smoothing GPS measurements, and GPS frequently updating the INS. This can result in very accurate and precise navigation.

The use of GPS systems has revolutionized the practice of surveying. Instruments and computer software to process the results have been developed specifically for various surveying applications. A review of current uses shows the broad range of applications:

- Rapid data collection is the goal of systems designed to be mounted in vehicles. Users include utility companies who survey the locations of their poles, contractors and highway departments who survey existing roads and planned routes, and governments who are converting their maps into computerized information systems. With these systems, the surveyor can drive from point to point, and by stopping for about 1 minute at each point and typing some information into the receiver, collect data for later computer processing. Data collection from several hundred locations can be done in one day, where previous systems required much longer times at each location, and teams of surveyors in order to measure between points. The resulting data can be processed into maps that have a accuracy of less than 1 meter.

- Surveyors use GPS receivers to survey locations that are located at a distance from the reference points they must use. In optical surveying methods, a direct line of sight had to be established, often by connecting points that went around or over obstructions like a mountain. Since each measurement required a team, one person at the last point and one at the next point, this could often be very time consuming and even dangerous in hazardous terrain. But with GPS, a receiver at the reference point and another at the desired end point provides precise results without regard to line of sight between them. And the precision is improved due to the elimination of several intermediate measurements.

- At the high-precision end of the surveying spectrum are the geodesists. These scientists measure the Earth to great precision for such purposes as tracking the movements of tectonic plates, or establishing major political boundaries. Geodesists often set up permanent GPS receiver sites and collect data continuously. One group that does this on a world-wide basis is the International GPS Service for Geodynamics (IGS). This group has a network of stations that include sites in North and South America, Europe, Asia, Africa, Australia and Antarctica, all of which collect data continuously. IGS collects the data and makes it available to anyone over the Internet world-wide computer network. Others have developed local clusters of stations that monitor areas of specific interest. For example, southern California contains an array of several hundred defined locations, some continuously occupied, to monitor the Earth movements that relate to the frequent earthquakes in that area. Receivers are used that are able to measure the locations with a repeatability of only a few millimeters over baselines that extend several hundred kilometers. When an earthquake occurs observations can determine which parts of the area actually moved, and in which directions and how far. This information permits the determination of where pieces of the Earth are divided, and which parts are connected together.

VIII. GPS Glossary

almanac
Information broadcast by all GPS satellites that gives approximate orbit parameters. Each satellite broadcasts almanacs for all satellites over a period of about 12.5 minutes. Almanacs are generally good for several weeks, and are used to predict which satellites will be visible at a specific time at a given site.

ambiguity resolution
The process of determining the integer number of carrier cycles between the receiver antenna and a satellite. The resolved integer number of cycles, added to the fractional part determined in the receiver, is used to convert carrier phase into a range measurement.

anti-spoofing (A-S)
A process by which the precise positioning signal (see P Code) from a GPS satellite is encrypted to preclude someone from sending a false signal that receivers might use in place of the real signal. The encrypted P Code is called Y Code. Decryption of the Y Code is reserved for users specifically authorized by the US Department of Defense. Anti-spoofing is referred to as A-S, which should not be confused with SA, or selective availability.

azimuth
The angular direction in the horizontal plane from a given reference point or direction (typically clockwise from North)

baseline
The line connecting two survey stations. In surveying, the error in a measurement is often expressed as some amount that specifies the actual error of the instruments, plus so many parts per million of the length of the baseline. This means that as baselines get longer, the total error in the measurement gets larger.

broadcast ephemeris
Orbit parameters (Keplerian elements) for a satellite as broadcast by that satellite. Used by the receiver and post-processing software to fix the location of the satellite in orbit at a specific time.

C/A code,
Coarse and Acquisition code
The Coarse and Acquisition signal modulated onto the L1 carrier of the GPS satellites. The code is modulated onto the carrier by 180° phase shifts of the carrier. The C/A code has a data transition, or chipping, rate of 1.023 MHz, thus providing a spatial bit length of about 293 meters. The C/A code is one of 36 unique Gold Codes with a repetition

interval of 1 msec, making them relatively easy to acquire. Sometimes called the Clear and Acquisition code since it is never encrypted. See P Code.

carrier phase

One of the observations made by a GPS receiver of each satellite. Once the receiver has a correct time reference, the distance to the satellite can be determined to be an integral number of carrier cycles plus a fractional part of a cycle. Using normal electronic techniques, the fractional part can usually be resolved to about 1 part in 1000. Determining the number of integer cycles is subject to some ambiguity, thus its determination is referred to as ambiguity resolution . Carrier phase resolution is referenced to the wavelength of the carrier, which for L1 is about 0.19 meters.

carrier wave

The basic signal sent by a radio transmitter. When a transmitter is specified as having a specific frequency, or channel, this refers to the frequency of the carrier signal. The signal varies over time from a high to a low level and back, thus having the shape of a sine wave. Since it travels through space at the speed of light, the distance between successive peaks is dependent on the frequency (number of cycles per second), and is referred to as the wavelength. As a point of reference, the carrier wave of the L1 signal from a GPS satellite has a frequency of 1572.42 MHz (or millions of cycles per second), and thus has a wavelength of 0.1903 meters.

chip, chipping rate

A chip is a single element of the binary data stream that is used to modulate a carrier wave in a spread-spectrum radio signal. The rate at which each chip appears on the signal is called the chipping rate. Since there is a frequency to the chipping rate, the signal containing the chips can be visualized as traveling in space just as the carrier wave does, and thus has an associated wavelength. In the case of the C/A code and P code on a GPS signal, the chipping rates are 1.023 MHz and 10.23 MHz, respectively. Associated chip wavelengths are approximately 293 meters for the C/A code, and 29.3 meters for the P code.

clock offset

The amount of time by which the receiver clock differs from GPS time. The clock offset is an unknown just as the receiver’s x, y and z position, and is solved for during the same position determination operation.

coherency

The phase relationship between two or more signals.

Signals that start and remain in constant relationship to each other, such as those that all start at 0, and return to 0 together after a fixed number of cycles of each signal, are said to be coherent. Signals whose relationship varies over time are said to be incoherent. Note that signals derived from a single signal source tend to be coherent, while those derived from separate sources that are not connected tend to be incoherent.

Control Segment
That portion of the GPS system that is responsible for monitoring the satellites' performance and sending control signals to the satellites. The main Control Segment facility is at Falcon AFB, located east of Colorado Springs, Colorado, USA. Other stations associated with Falcon AFB are located around the world.

data link
A communications channel which permits the transfer of information between two devices. Generally, this refers to a digital data transfer. Specific examples include a pair of radios, two telephones, and cables.

data word
The smallest unit of the GPS navigation message. Each data word consists of 30 bits (or binary digits - 1s or 0s), of which the first 24 bits are actual data, and the last 6 are referred to as parity, and are computed from the previous bits so that a receiving system can ensure that the data received was correct. A total of 10 data words together constitutes a subframe.

dataBit checking
A technique where data received in the navigation message (ephemeris, almanacs) are tested for reasonableness, and discarded by the receiver if they are found to define unreasonable orbits (such as orbits that would be below the surface of the earth, or beyond the moon).

Differential GPS, DGPS
An arrangement of receivers in which one acts as a reference station, and others act as user stations. The reference station observes satellites and computes correction factors for the others. The user stations apply the corrections to their own observations and generally obtain position solutions that are improved by a factor of 10 or more over C/A code solutions.

Dilution of Precision (DOP)
A factor which allows the amount of error in measurements made from the GPS satellites to be translated into a measure of the error in the resulting position. There are several

specific DOPs which are defined elsewhere in this glossary, including GDOP, HDOP, PDOP, TDOP and VDOP. In general, DOPs are derived from the diagonal of the design matrix used to determine position and clock error, and depend on the number of satellites present (more satellites yield better DOPs) and their relative locations (greater geometric dispersion yields better DOPs). In use, the position accuracy may be computed by multiplying the measurement accuracy by the associated DOP. A DOP value ranges from 1.0 (the best possible) upwards.

double difference solution A GPS solution where data from two satellites, both observed simultaneously by two different receivers, are differenced in such a way that clock errors in both satellites and both receivers are removed.

ECEF coordinates Earth Centered, Earth Fixed coordinates. Coordinates used by the GPS system, implying a Cartesian coordinate system referenced to and fixed by the Earth. In general terms, the origin is at the Earth's mass center, with the z axis along the Earth's mean rotational axis, positive to the north. The positive x axis emerges from the Earth at 0° N, 0° E, and the y axis forms a right-hand coordinate system by emerging at 0° N, 90° E. This coordinate system is that defined by WGS84 .

elevation angle The angle at an observation point between a local horizontal plane and an object. 0° is horizontal, positive angles are above the horizon, and 90° elevation corresponds to the object being directly overhead. Complement of zenith distance.

elevation mask (angle) The elevation angle below which satellites will not be acquired by a GPS receiver. In session planning, this angle is used to limit when satellite information is shown in the displays. In a receiver, it is used to limit when satellites will first be searched for (rising satellites), and in some instances when the satellite will be dropped from further tracking (setting satellites). Since satellites close to the horizon are frequently either masked by obstructions, or subject to observation errors due to signals arriving at the receivers from multiple pathways (see multipath), the elevation mask is set to exclude lower satellites. In typical surveying applications an angle of 15° is used. Also called horizon mask, since it effectively masks the local horizon.

ellipsoidal height The vertical distance between a point and the surface of the reference ellipsoid.

ephemeris The set of parameters that together describe a satellite's orbit, and where the satellite is in that orbit at a given time. As used in GPS, the ephemeris consists of Keplerian elements that describe the orbit's shape, its orientation with respect to the stars, and its position in the orbit with respect to a stated reference time. The ephemeris is used to compute the location of the satellite at a particular time of interest with respect to the Earth. Note: plural of ephemeris is ephemerides (pronounced "ef-em-air'-i-deez").

epoch 1) A sample interval in a GPS receiver. 2) A GPS week. 3) An instant in time.

frame The standard unit of data in the GPS navigation message. The frame consists of 5 subframes containing a total of 1,500 bits of information, and requires 30 seconds to transmit at 50 bits per second.

GDOP Geometric Dilution Of Precision, a measure of the error in position and clock in a GPS solution. GDOP is composed of both position and time DOPs, where $GDOP^2 = PDOP^2 + TDOP^2$, and is considered the all-encompassing DOP value. All other DOP values are components of GDOP.

geodetic survey A survey taking into consideration the true shape of the Earth usually done to a higher precision or accuracy and normally extending over greater areas than a normal land survey. Geodetic surveys are generally done for such purposes as defining the starting point for future land surveys, or monitoring the changes in the shape or configuration of the Earth (as in monitoring land changes due to earthquakes).

geodesist A scientist who specializes in performing geodetic surveys or in interpreting their results.

Gold Code The type of binary data stream used to modulate the GPS L1 carrier with the C/A code. The GPS Gold Codes are 1,023 bits long, and have approximately the same number of 1s and 0s. The Gold Codes can be generated by two digital shift registers, each with specified feedback connections, whose outputs are combined by logical exclusive OR functions. All possible Gold Codes may be generated by the same two shift register setups. To create each Gold Code, the output of one shift register is delayed by a set number of bits from that of the other before the two outputs are combined.

GPS observation The data collected by a GPS receiver based on observations of individual satellites. Observations consist of such elements as pseudoranges computed from the C/A Code and/or P Code, carrier phase measurements on L1 and/or L2 carriers, and error estimates of these values.

GPS Time, GPS Week The time used in the GPS system is derived from a composite of all of the atomic clocks in the satellites and Earthbound monitoring stations. It is steered to be within about 1 microsecond of UTC with respect to the start of each second. GPS time is kept and reported as an integer number of seconds since the start of January 6, 1980. Since GPS time does not add any leap seconds, it differs from UTC by the integer number of leap seconds added to UTC since January, 1980 (twelve, as of July, 1997). GPS weeks start with the first second after midnight (GPS time) on Sunday of each week, with the week of January 6, 1980, referred to as week 0.

HDOP Horizontal Dilution Of Precision, a measure of the amount of error in GPS position determinations in the local horizontal plane. HDOP is chiefly influenced by the number and relative geometry of the available satellites. Since satellites can generally exist in all horizontal directions from the receiver, HDOP values are usually less than VDOP values, which are limited by the inability of the receiver to see satellites located below the local horizon. See Dilution Of Precision and PDOP.

height (ellipsoidal) The distance of a point above a reference surface. In GPS measurements, height is normally computed above the WGS84 reference ellipsoid.

horizon mask See elevation mask.

Inertial Navigation System (INS) A navigation system that determines where it is located by starting from an initial entered position, and then keeping track of all movements of the vehicle. Changes in attitude of the vehicle are sensed by gyroscopes, and changes in velocity are sensed by accelerometers. An INS is generally very sensitive to immediate vehicle movements, but is subject to errors that accumulate over time since the last correct position input.

integer bias The ambiguity in the number of whole carrier wave cycles between the receiver antenna and the satellite. Resolved by ambiguity resolution .

ionosphere That region above the Earth that contains ionized particles that can affect the propagation of radio signals. The effect on signals varies with the amount of free charges, which are created generally by solar radiation, thus exhibiting a large diurnal variation, and a smaller variation dependent on the solar cycle. Since the effect varies with frequency, the transmission of two signals with different frequencies allows dual-frequency receivers to determine the effect and correct for it. The effect on a GPS signal is to advance the carrier phase, but to delay the modulation codes (C/A and P codes). The thickness of the ionosphere varies between day and night, and over the course of the 11-year sunspot cycle. It generally exists between 50 and 1000 km above the Earth's surface.

Keplerian elements A set of parameters that describe the orbit of one body about another. There are seven basic Keplerian elements which together describe the shape of the orbital ellipse (length of the semi-major axis and eccentricity), the orientation of the ellipse with respect to the Earth (longitude of the ascending node, inclination angle of the orbit, and argument of perigee), and the location of the satellite in the orbit at some specified time (mean anomaly and mean motion difference). In addition to these seven terms, the GPS system also broadcasts other terms that describe how the orbit changes over time. Used together, all of these terms allow determination of a satellite's position in space at a specific time to within about 0.3 meters (1 foot).

Kinematic A receiver mode that makes continuous measurements while the receiver may be moving. Ambiguity resolution is done as a continuous process, and each observation yields an independent point position.

L1, L2 The carrier frequencies transmitted by the GPS satellites. L1, the primary signal, is transmitted at 1575.42 MHz, and is modulated by spread-spectrum techniques (phase transitions at specified "chipping rates") with the C/A Code and P (or Y) Code. L2, the secondary signal, is transmitted at 1227.60 MHz, and is modulated only by the P (or Y) Code. Each signal is modulated with a Pseudo-Random Number (see PRN identifier). The detection of this PRN identifier, possible even in the presence of several signals, is what allows all satellites to broadcast on the same frequencies. The L2 signal is broadcast in addition to the L1 in order to allow receivers which receive both frequencies to determine how much the ionosphere has affected the signals. Note that later satellites may add other codes onto L1 or L2, and may add additional carrier frequencies.

model range The distance between a satellite and receiver as computed using the known position of the receiver, the broadcast ephemeris of the satellite, and the current time.

multipath A GPS error source. When signals from the satellite reflect off of objects such as buildings, terrain features or other objects, they travel in longer paths and arrive at a later time than the signal arriving by direct path. The effect is to interfere with the direct signal and make proper determination of the correct signal less precise.

navigation message The data added onto the GPS PRN signal. It contains such data as the current GPS time in seconds, correction terms for the satellite's clock, the broadcast ephemeris, and almanac data for other satellites. The navigation message consists of frames, each composed of 5 subframes, each of which is composed of 10 data words which are 30 bits long.

P code, Precise code The precise position signal sent by the GPS satellites. P Code is added to both the L1 and L2 carriers by phase shifts at a chipping rate of 10.23 MHz (equivalent to a spatial bit length of about 29 meters). When anti-spoofing is turned on, the P Code is encrypted and is then referred to as Y Code.

PDOP Position Dilution of Precision, a measure of the error contained in the 3-dimensional position determinations made by the receiver. PDOP is determined from the number of satellites and their relative geometry. PDOP is related to HDOP and VDOP by: $PDOP^2 = HDOP^2 + VDOP^2$. PDOP and TDOP are the two orthogonal components that comprise GDOP. See Dilution Of Precision and the other DOP entries.

phase
See carrier phase.

phase center, phase-center offset
That location in an antenna at which the signal appears to be received. While the entire antenna actually participates in receipt of the signal, when positions are computed using very precise setups, a single point can be computed where the signal appears to be detected. While this point may vary somewhat with angle of arrival of the signal, and with frequency of the signal, it may be computed to a precision of fractions of a millimeter for given conditions. The location of the phase center is important in high-precision surveys since it represents the point in space to which point-position results refer. The phase-center offset is the difference between the phase center and the mechanical center of the antenna.

point positioning
The process by which observations made at a station are combined to compute the location of that point relative to the reference system. In GPS work, the GPS observations which determine a range between the satellite and the receiving antenna are combined with the position of the satellite at the time of the measurement, as computed from the ephemeris. Each combination of range and satellite position defines a sphere. All spheres thus defined are then increased or decreased uniformly until a single point (or minimum volume) is defined. The coordinates of the point are the x, y and z position of the receiving antenna, and the uniform amount by which the spheres were adjusted is the receiver's clock offset from GPS time.

precise ephemeris
Ephemeris for a satellite generally computed from external measurements (such as from several fixed locations, or from laser or optical observations) which can be used to define the satellite's orbit and location more precisely than the broadcast ephemeris. Precise ephemerides are computed by several groups (e.g., US National Oceanographic and Atmospheric Administration, International GPS Service for Geodynamics) and may be posted on the Internet for use in high-precision post processing.

PRN identifier
Pseudo-Random Number identifier. This is the basic modulation pattern applied to the carriers of each satellite. The satellites are each modulated with a modulation pattern generated by a hardware pseudorandom number generator seeded with a starting value referred to as the PRN

identifier. For the C/A code, there are 36 possible Gold Codes, and each is unique over the 1 msec interval before it is repeated. Since each satellite uses a different PRN identifier, the satellites are typically referred to by this value. The GPS satellites only use 32 of the possible 36 PRN identifiers, since only 5 data bits are allocated in the data the satellites broadcast to specify each satellite. The remaining codes are reserved for other uses, such as ground transmitters.

pseudorange The computed distance to a satellite derived from the time that the signal took to arrive at the receiver. Pseudorange is determined from the C/A Code and/or P Code, and is thus limited in resolution by the spatial bit length of the codes. Compare to carrier phase.

RAIM Receiver Autonomous Integrity Monitoring - a software process that computes position solutions using all available satellites by taking them in groups of the minimum necessary. In each solution, the error is computed and compared with that from other solutions looking for any specific satellite that is contributing excessively to the error. If such a satellite is found, it is generally marked as bad and not included in final solutions until such time as it is determined to have improved its performance.

rapid static A receiver mode similar to static but with predefined error limits larger than those for a static mode measurement. The larger error limits permit data collection to those limits to be completed in a shorter period of time than that required to attain the limits specified for static. The receiver is not required to track satellites between points, and may be shut off.

Real-Time Turbo Kinematic, RTTK A form of differential GPS in which information about the carrier phase is broadcast by a reference receiver, permitting another receiver to perform ambiguity resolution and determine position ultimately to fractions of a meter in real time. The remote receiver uses double-difference calculations to remove satellite and receiver clock errors, thus permitting the higher precision.

reference station, reference receiver A GPS receiver located at a fixed point during the entire time that observations are made at other stations using other receivers. The observations made by the reference station may be used to provide double difference solutions for the other receivers, thus refining the observations to improve the solutions. The term also applies to the receiver that generates corrections for use by other receivers in a Differential GPS operation.

right-hand coordinate system
An x, y, z coordinate system in which the directions of the positive axes may be represented by the thumb, first and middle fingers of the right hand positioned to be at right angles to each other. The thumb and extended first finger held at right angles to each other in the plane of the palm represent the +x and +y axes, respectively. The middle finger, bent to extend perpendicular to the plane of the palm, represents the +z axis.

RINEX
Receiver INdependent EXchange format, a file format adopted by the International Association of Geodesy (IAG) for standardizing the data output of GPS receivers. Most manufacturers provide a means to either produce data in this format, or to translate their own format into RINEX. Most commercial post-processing software packages accept this format for input data.

sampling interval
The period of time during which observations of the GPS satellites are collected and combined into a single reported value. Typical sampling intervals range from 1 to several hundred seconds, and typically consist of the accumulation of from 1 to 50 observations made each second (depending on the specific receiver).

selective availability (SA)
A process by which the C/A Code signal of the GPS satellites is altered to produce errors in the computed positions of receivers. The signals are altered by such means as giving inexact ephemerides or clock correction factors, or manipulating the satellite's clock. When SA is activated, position solutions generally exhibit errors of up to 100 meters 2D RMS. Do not confuse SA with A-S (anti-spoofing).

Space Segment
That portion of the GPS system consisting of the satellites in orbit. A complete satellite constellation is defined as 24 satellites arrayed as 4 each in 6 separate orbits. The orbits are inclined about 55° from the plane of the equator, and are evenly spaced at 60° intervals about the equator. The satellites are generally evenly spaced within the orbits, thus

about 90° apart from each other.

spread spectrum A method of transmitting a radio signal by sending a pure sine-wave carrier at a fixed frequency, and periodically altering the phase of that signal at a rate defined as the chipping rate. The pattern of phase changes is defined by a binary bit stream that appears to be random, but is in fact generated by a process that makes it predictable to someone who knows the process. Detection of the signal is performed by generating a matching chip pattern, and sending the received signal and the matching pattern into a circuit called a correlator. When the chip pattern is present and in time coincidence with the matching pattern, the correlator will generate an output that can be recognized by other circuits. The term spread spectrum refers to the effect of the periodic phase changes causing the signal to appear to be spread out over a range of frequencies rather than existing as a single sharp spike at the carrier frequency. Note that other forms of spread-spectrum also exist, such as periodically changing the transmitting frequency (“frequency hopping).

static A receiver mode in which the receiver is positioned over the survey station and left to collect observations. No provisions are made to accommodate movement of the receiver between points, but rather each such survey is assumed to be an independent set of observations. Site occupation can range from several minutes to several days. This is the most accurate type of GPS survey. Sometimes called Fixed.

station A location that is to be surveyed.

stop and go A receiver mode in which the receiver continuously tracks satellites while it is moved from station to station. At the station, some action (such as pressing a button or entering data) is taken to tell the receiver when the survey of a station starts and ends. Site occupation is typically reduced to a few epochs since the ambiguity resolution is maintained by the receiver during the movement process between stations.

subframe A unit of data in the GPS navigation message. Each subframe is composed of 10 data words, each of which are 30 bits long. Thus, the subframe is 300 bits long, and requires 6 seconds to transmit at 50 bits per second. There are five defined subframes, which when broadcast together in sequence comprise a frame.

TDOP Time Dilution Of Precision, a measure of how much error exists in the time solution. Unlike other DOPs, TDOP does not vary with location of the satellites, but only with the number of satellites used. TDOP and PDOP together comprise the orthogonal components of GDOP. See Dilution Of Precision, GDOP and PDOP.

troposphere That layer of the Earth's atmosphere closest to the Earth's surface. The troposphere is where the majority of the weather occurs. Its thickness varies from somewhat less than 9 km at the poles to over 16 km at the equator.

UERE User Equivalent Range Error, a measure of the total error in a pseudorange measurement, including such factors as propagation errors, ephemeris uncertainties, clock and timing errors, and noise.

user receiver A receiver used in a differential GPS setup (either in real-time or for post processing) that makes use of the data generated at a reference receiver to compute position. Sometimes called a rover receiver.

User Segment That portion of the GPS system composed of all of the receivers currently listening to the Space Segment for purposes of determining pseudoranges and computing positions. There is no specified organization to the User Segment, but rather it consists of whichever receivers might be present at any one time.

UTC Universal Time Coordinated, the uniform atomic time maintained in the US by the United States Naval Observatory, and used as the basis of GPS Time. Note: GPS time differs from UTC time by the number of leap seconds that have been added to UTC since January 6, 1980, and by a variable offset usually kept to less than 1 microsecond.

VDOP Vertical Dilution Of Precision, a measure of how much error exists in height determinations due to the geometry of the satellites used. VDOP is most influenced by the dispersion of satellites in elevation. Since there is a physical limit on satellite visibility below the horizon, VDOP values are generally larger than the associated HDOP. VDOP and

HDOP are the orthogonal components of PDOP. See Dilution of Precision, HDOP and PDOP.

WGS84 Ellipsoid

World Geodetic System 1984, the standard datum used by the GPS system. For this system, the Earth is defined by an ellipsoid centered on the computed mass center of the Earth. Viewed from above the north pole, the ellipsoid appears circular. Viewed from above the equator, the ellipsoid looks like an ellipse that is wider than it is high. The semi-major axis (distance from the center of the Earth to the equator) is set at 6378137.0 meters, with an inverse flattening value (flattening is defined as the difference between the semi-major and semi-minor axes divided by the length of the semi-major axis) of 298.257223563.

Y Code

The encrypted form of the P Code (see P Code).