GPS explained

EpiphanyBySteveLee.com

Table of Contents:

Introduction:

The Global Positioning System (GPS) is a US DOD radio navigation system that employees 24 non-geosynchronous satellites (plus spares). These satellites continuously broadcast time-stamped signals that are used by receivers to calculate their location on or near the surface of the earth. GPS is a passive system, meaning that the user equipment does not need to transmit any information to resolve its location. The receiver does however need an unobstructed view of the sky in order to detect signals from at least four GPS satellites. Operating a GPS receiver in an area with many tall reflective surfaces (e.g. nearby tall buildings, granite mountains, etc.), and/or dense overhead canopy can introduce errors in the location solution, particularly when the receiver is forced to use a GPS signal reflected off one of those tall reflective surfaces. In these cases the location estimate can be off by as much as several hundred meters/yards or more.

Generally speaking, the accuracy of the location estimate is affected by a number of things, including the distribution of the visible satellites overhead across the sky, the density of the ionosphere, and the quality and state of the user's GPS receiver (etc.). Most consumer grade receivers typically offer an accuracy within ± 2-20 meters/yards, while precision or military grade equipment can resolve the location down to centimeters/inches or less.

System overview:

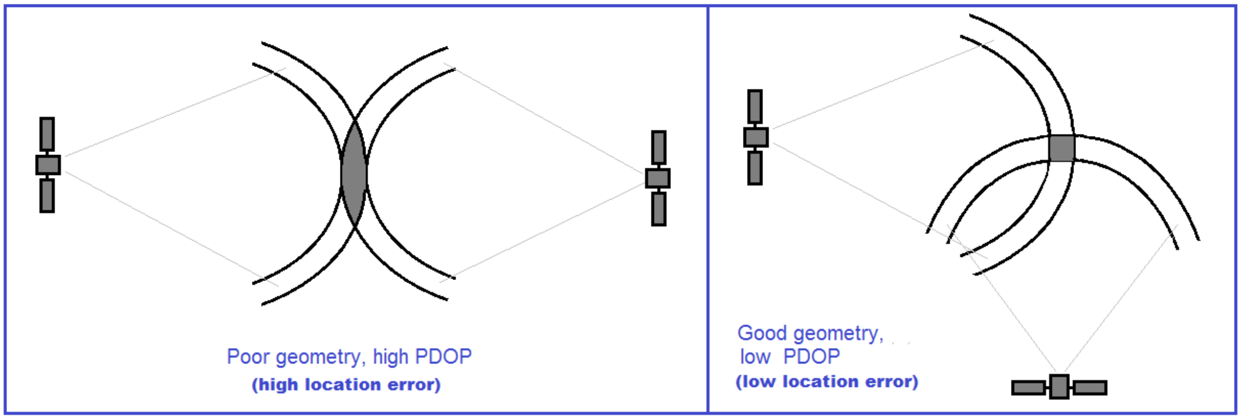
GPS satellites orbit the Earth approximately 12,600 miles (20,000 km) up, at a rate of one complete orbit every 12 hours. Each satellite vehicle ("SatV") transmits time from several high-precision on-board atomic clocks, as well as information related to the satellites' trajectory/location (i.e a table known as the "ephemeris", from the Greek word meaning "short lived"), as well as SatV health/status, ionospheric modeling estimations, etc. (in what is known as the "almanac" portion of the signal). This information is updated periodically by a network of ground-based control stations and transmitted up to the satellites. At the SatVs this information is then transmitted down to the users using *Direct Sequence Spread Spectrum Code Division Multiple Access* or "DS-SS CDMA" (this is very similar to the method used by 2G and 3G CDMA cell phones, though without the use of orthogonal Walsh Codes; see this authors other tutorial on cell phones at "epiphanyBySteveLee.com").

Finding one's location on the Earth using GPS is all about measuring the "time of flight" of the GPS radio signals (i.e. the time a signal takes to travel from the SatV to the user). To determine the "time of flight", the GPS receiver "simply" reads the time stamp in the SatV transmitted messages and then subtracts that time from its local clock time.

Once the receiver knows the "time of flight" it can convert that into the distance from the SatV (known as the "pseudo-range") using the simple relationship of "distance = rate x time". Now that it knows how far it is from this SatV, it can use the ephemeris data to estimate where that SatV is relative to known reference points on the Earth, and then combine that location with the pseudo-range from the SatV to calculate a rough estimate of its location on or above the surface of the Earth. This pseudo-range only gives its distance from that one SatV, without any directional information. Knowing only the distance from this SatV therefore effectively scribes out a large circle or oval over the surface of the Earth, placing the user somewhere along that circle/oval.

To narrow that location estimate down, we must generate another pseudo-range circle relative to another SatV. Since we must be on both circles (i.e. we must be "x" km from the first SatV and at the same time also be "y" km from the second SatV), that places us somewhere in the area where these two circles overlap, reducing our possible locations down from "anywhere on this one large circle" to only the area where these two pseudo-range circles overlap (somewhat analogous to a Venn diagram).

Generally speaking, anytime we want to define a specific location of something in *four*-dimensional space (x,y,z, and time), we must define four variables (x,y,z,t). For example if you want to track a bouncing ball, you will need to specify not only its "x" and "y" location relative to some reference point, but also its height "z" – while at the same time being specific about where it is each moment in time (thus x,y,z,and t). To get a good four-dimensional location estimate, our GPS receiver must lock onto and generate pseudo-range values from at least four different SatVs (i.e. since we have four variables, we need four simultaneous equations in order to solve for each of those variables).



<u>"Dilution Of Precision" (DOP)</u>:

From the figure we can see that the geometrical distribution of the SatVs overhead used in the solution has a direct impact on the accuracy of the location estimate (i.e. how big of an area of overlap in the pseudo-range circles). This geometry-related error is described in GPS literature using a term known as "PDOP", or "Positional Dilution Of Precision" (DOP is a unit-less multiplier in the position solution). In order to calculate the highest accuracy location solution possible, we need to reduce the amount of overlap between all of these pseudo-range circles. In other words, since the estimate of our location effectively equates to overlap of all of these pseudo-range "beam footprints" in the area around the receiver, we want the area of that overlap to be as small as possible.

If for example, the SatV signals used in the location calculations all come from the same area of the sky (which can occur for example when the receiver is in a long, narrow "urban canyon", or otherwise has limited view of the sky, etc.), we end up with a great deal of "redundant overlapping" in our pseudo-range footprints, resulting in a very large overlapping area which implies a much higher ambiguity to our location estimate. If on the other hand all of the signals coming from the SatV used in our location calculations come from very different regions of the sky, there will be much less "redundant" pseudo-range overlap, and therefore a much more reduced area for our location estimate.

Generally speaking, a PDOP value of less than three provides more accurate location estimates, while PDOP values greater than seven provide poorer location accuracy. Since the trajectory of each SatV is well known (e.g by DOD tracking stations), it is easy enough to predict the amount of PDOP will exist for a given day and hour at any specific area on the globe, allowing users to plan their GPS usage to coincide with the lowest PDOP and thus highest location accuracy (for those interested, a number of good PDOP planning tools are available on-line, including one at www.Trimble.com),

Cold vs. Hot starts:

Each receiver starts its location estimation process by using an estimated position of each SatV (which it calculates using the daily ephemeris table) to help it first “hunt for” and then lock (“sync”) onto each of the very weak SatV signal. Once a receiver locks on to the broadcast signal, it then measures the “time of flight” for each GPS satellite signal, i.e. the difference between the user’s receiver clock time and the time stamp in each GPS SatV transmitted data stream.

The amount of time it takes a receiver to “lock” on to a SatV signal depends on how long it has been since it last locked onto (and read the ephemeris data from) a GPS SatV. A “cold” start implies it has been more than four days since the receiver last locked to any GPS SatVs, and therefore does not have a current ephemeris or current system clock sync. A “warm” start implies the ephemeris is less than four days old but the on-board clock hasn’t sync’ed with a SatV within the past four hours (which requires the receiver to work a little harder in order to get sync'd up to the GPS system timing). A “hot” start implies the receiver has a recent ephemeris and recent clock sync (i.e. within the past few hours). Receiver SatV “lock” can take ~30 minutes for "cold" start, ~2 minutes for a "warm" start, and ~2 seconds for a "hot" start. (Note that some “network connected” devices with GPS chips built in, such as a cell phone, laptop/wearable WiFi equipped device, etc. can lock on much quicker even if it has been weeks since last receiving data from the GPS SatVs, if the network connection “assists” it by feeding it a current ephemeris, a technique referred to as “network assisted GPS”).

Error sources:

On paper this whole process sounds fairly straight forward: to find our location, we “simply” determine the “time of flight” of the GPS radio signals using our GPS receiver's clock, plug it into the “distance = speed x time” equation to generate the pseudo-range distances from each SatV, and then use each SatV's estimated location (derived from the ephemeris) to scribe out some pseudo-range circles on the Earth, and where these circles overlap is where we must be.

In practice however a number of real-world imperfections in our process limit the accuracy of our location solution.

Perhaps the most obvious flaw in our naive perception of things is the fact that everything in this scenario hinges on our ability to *accurately* measure time, and therefore on our clocks ability to keep *perfect* time. When it comes to time and spatial locations however, “there are no absolutes”, or as physicists like to say, “everything is relative”. Stated another way, there is no such thing as a perfect clock. Even if you buy a *very* expensive time piece and set it “exactly” to some “perfect” reference (e.g. WWV in Fort Collins, Colorado), we all know clocks drift over time and at different rates, depending on the temperature, humidity, planetary alignments, mood, etc. The good news is that the clocks used in the SatVs are of much higher quality than your average consumer-grade clocks which helps, plus they are constantly updated as needed by the ground-based control stations. However the same can not be said of the clocks in the average GPS receiver. Consequently they tend to not only suffer a fair amount of drift over time, they also experience a certain amount of “clock jitter” which only complicates their ability to lock onto the signals. Consequently, when the typical consumer-grade receiver attempts to calculate “time of flight” by subtracting the received SatV time from its own [inaccurate] clock time, it can grossly over- or under-estimate the pseudo-range distance. And even though the SatV is telling the receiver what the “true” system time is, the receiver initially has no way of knowing how “old” that time is (i.e. the total “time of flight”). Even one *millionth* of a second in uncertainty/error corresponds to 300 meters in location error (since $D = V \times T = 300 \times 10^{6}$ m/s $\times$ 10^{-6} seconds = 300 m). (Recall that time is one of the four unknowns in our location solution, i.e. x,y,z,t; to determine the GPS system time, the receiver must go through the solution process, locking onto and calculating pseudo-range estimates from at least four SatVs.).

If all of this wasn't enough of a flaw in the best made plans of mice and men, we note that in addition to the imperfection in our hardware, there are a number of atmospheric quantities that fluctuate over time, affecting the speed of those GPS radio signals as they travel through the atmosphere. These atmospheric fluctuations in speed create some additional amount of uncertainty in our “time of flight” calculations, which then translates into additional uncertainties in our pseudo-range location estimate.

When we add the uncertainties due to the imperfections in our hardware (e.g. clock drift, clock jitter, etc.) with the uncertainties in the speed of propagation used in our “distance = speed x time” calculations, our pseudo-range values end up with a significant amount of uncertainty (or “fuzziness”) in them, making this range now ”x” km ± *some uncertainty*. This “± some uncertainty” in the arc of each SatV pseudo-range circle effectively “fattens” it out from a thin line into a broad *swath* that represents the “fuzziness” introduced by the uncertainty of these varying factors. This broad swath in our pseudo-range then translates into a location error on the order of several meters/yards or more. This is all in addition to the location error related to PDOP.

Fortunately most of the larger variations on the “time of flight” due to changes in the atmosphere can now be modeled fairly accurately, and their impact on the “time of flight” calculations compensated for during the pseudo-range calculations (using the “almanac” information broadcast by the SatVs). These variations include changes in the moisture content in the atmosphere due to daily and seasonal weather changes, as well as fluctuations in the ion density in the ionosphere due to solar radiation, solar “winds” (ions spewed out by solar flares), etc. There are however other less predictable fluctuations, particularly in the ionosphere that are much more difficult to model and thus much more difficult to eliminate in the location calculations. These fluctuating factors add an additional several meters or more to the uncertainty in the location solution. (See additional Ionosphere related discussion below.)

In practice, these solution accuracy factors are typically described as two separate but related "components". The first is related to the geometry of the satellites used for a solution, quantified by the previously mentioned "Dilution of Precision" (DOP). The second "component" in the solution error is the GPS satellites' User Range Error (URE).

From this discussion, we see that there are several different types of URE. SatV position estimate and clock estimate errors are known as Signal-In-Space UREs (SISURE) because they deal with satellite-based errors. Other UREs contain errors that arise from the user receiver's processing, atmospheric modeling errors etc. All of these URE sources accumulate and add to the final position error. SISUREs are not very predictable, many being the result of quantum fluctuations of the on-board atomic clocks on the GPS satellites, transient ionospheric perturbations, etc. This random behavior means that the quality of the position solution is inherently affected by the inescapable randomness of quantum objects (e.g. in the atoms/molecules in the crystal oscillator, ionospheric electron densities, etc.).

In short, position error is a combination of a predictable geometrical value (PDOP), and an unpredictable range error related to each satellite being tracked. Consequently the typical overall position error experienced by the average consumer-grade GPS receiver can be on the order of 2-20 meters. High precision receiver$ can generate solutions with an accuracy at half a meter to a few centimeters (or better in some case$).

Correction factors:

The "time of flight" path distance is referred to as a "*pseudo*-range" ("rho", or "ρ") due to these several varying factors (e.g. clock timing error, weather related atmospheric moisture fluctuations, variations in ionospheric propagation delays due to fluctuations in ion densities, etc.). The GPS receiver can correct for some of these errors, but to do so, it needs to obtain certain corrective factors related to the current atmospheric conditions at that time. These corrective factors are provided by the GPS system in the almanac that is periodically broadcast by the system in tandem to the ephemeris. Once a GPS receiver has established its initial lock It takes roughly 30 minutes of continuous reception for it to acquire the most current almanac and ephemeris information from the SatV's. (Technical details for building a receiver and using the data from a GPS satellite are contained in the IS-GPS-200D).

Once the receiver has all this corrective information from the satellites, it applies the corrected pseudo-range measurement, "ρ_c", in a linearized least-squares algorithm to generate a more accurate position solution. This equation provides an estimate of position, using an initial starting position "guess" (typically ball-park at best), which it then improves upon in subsequent iterations, using the position estimate result from the previous calculation as the initial estimate for the next calculation. These iterations then continue to "hone in" on the best location estimate until the change in the position estimate falls below some minimal threshold, at which point the receiver reports a valid lock has been established.

The transmitted GPS signal:

As for the actual signals being broadcast by GPS satellites, the system in operation today actually uses two different “channels” or frequencies to transmit its information: L1 (1575.42 MHz) and L2 (1227.60 MHz). The time-stamped data (including the previously mentioned ephemeris and almanac) running at 50 b/s are used to “modulate” two “Pseudo-Random Noise” (“PRN” or simply “PN”) codes, spreading the energy of the 50 b/s data stream over the full transmit channel in the process. This spreading technique using the PRN codes not only enables the user receiver to identify which SatV it is hearing, but as a byproduct of the spreading process, creates a “spreading gain” that allows the receiver to extract the weak signals out of the noise. These spread signals are then used to modulate the L1 and L2 transmit carriers, which product is then amplified and sent out the SatV antenna. (FYI - The atomic clocks in the SatVs generate a fundamental 10.23 MHz base clock signal, which is then multiplied up to the carrier frequency by 154 for L1 and by 120 for the L2 carrier.)

In the current GPS system, there are actually two codes used during modulation: the Coarse Acquisition code (“C/A”, using what is known as a “Gold code”), and military-only Precision (“P”, or if encrypted “Y”) code. P/Y-code is transmitted on both L1 and L2, while C/A-code is carried only on L1 (1575.42 MHz). Since the speed of all electromagnetic waves is *frequency dependent*, transmitting the P/Y code on two different frequencies (L1 and L2) helps the military grade receivers isolate, and thus compensate for some of the “time of flight” uncertainties, enhancing the accuracy of their location solutions. Newer SatV’s are being equipped with a third channel designated L5 and transmitting at 1176.45 MHz, though it is not expected to be fully operational for several years (i.e. when sufficient number of SatV’s and ground receivers are equipped with L5). The “chip rate” or spreading rate for C/A code is 1.023 Mb/s, while the chip rate for P code is 10.23 Mb/s. The C/A code repeats once every 1mS, while the P code repeats once per week. Both codes are re-initialized at midnight between Saturday and Sunday.

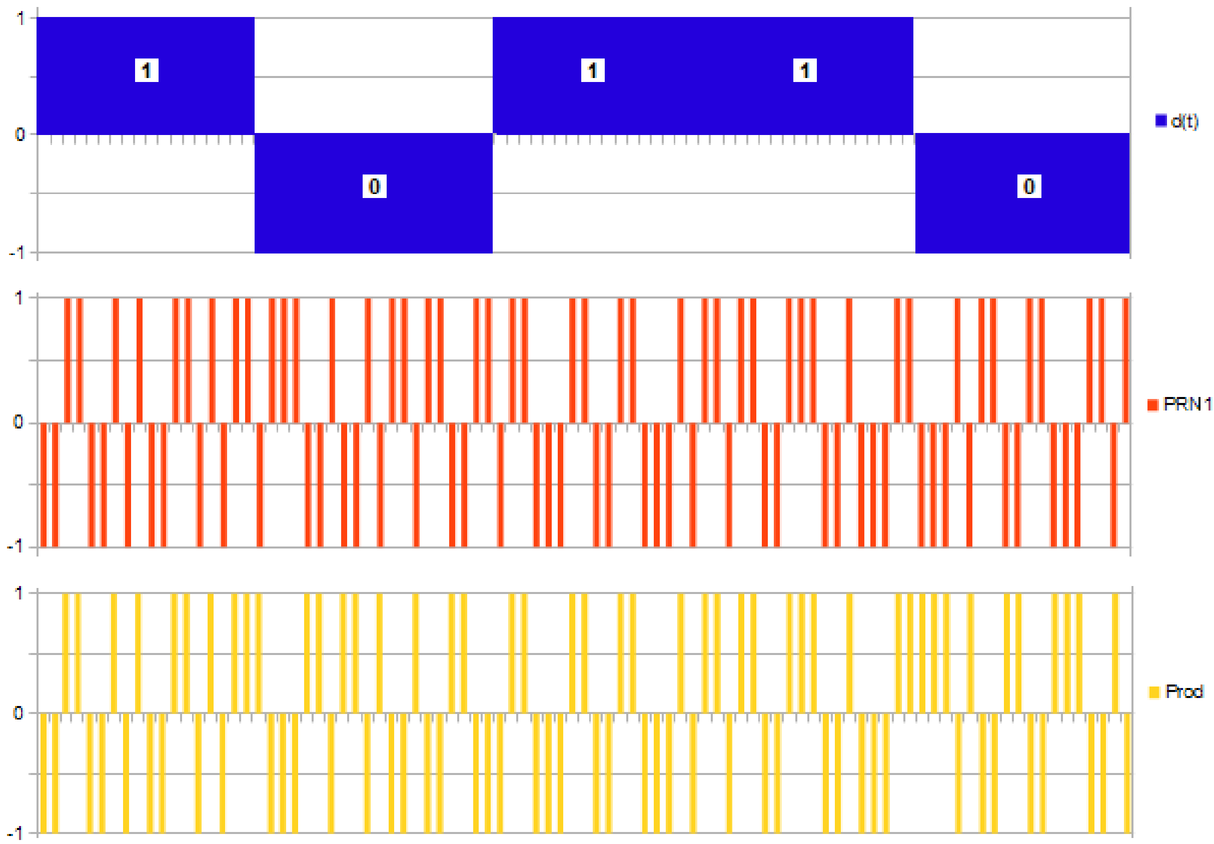
Spreading gain “G_S”:

Since the SatVs are orbiting 20,000 km above the Earth and each SatV has only a limited power supply (i.e. limited transmit power), by the time the signals arrive at the user, the signals are extremely weak (~-155 to -160 dBm[1], which is well below the noise floor of the typical receiver). By spreading and then despreading the bit stream across

1 Where “dB” = “deciBells” (deci = 1/10th , dB = 1/10th of a Bell, where a Bell is a power of ten; e.g. 20dB = 10^2). dBm = decibels relative to 1 mWatt, thus dBm = 10 x Log (Prx / 1 mWatt).

the channel using a DS-SS CDMA process, the system achieves a “spreading gain” (“G_S”) that allows the very weak GPS signals to be “pulled out of the grass”.

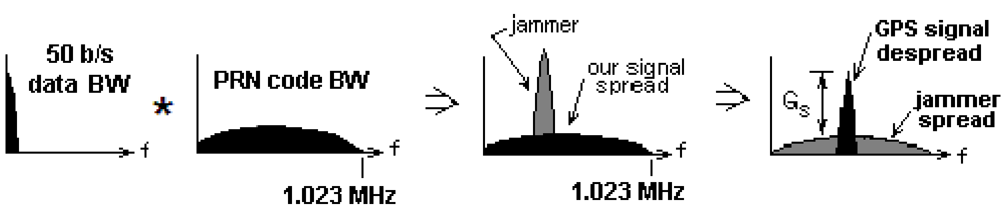
To understand how this works, we note that when the 50 b/s data stream is transmitted, it is convolved against the much higher rate PRN codes, as depicted in the figure below (not to scale) (this can almost be thought of as “modulating” the PRN code). The essential aspect of any such PRN convolution is that the “daughter product” of the process effectively inherits the characteristics of both its “parents”. This means that the resulting product has all of the information that was in the 50 b/s data stream, while at the same time has taken on the much higher bit rate of the PRN code as well, along with the PRN's well-defined bit pattern. This effectively takes all the energy that is in the 50 b/s data stream (“d(t)” in the figure) and spreads it out across the wider 1.023 MHz or 10.23 MHz PRN bandwidth. This product is then used to modulate the main L1 and L2 carriers and then transmitted down towards the surface of the Earth.

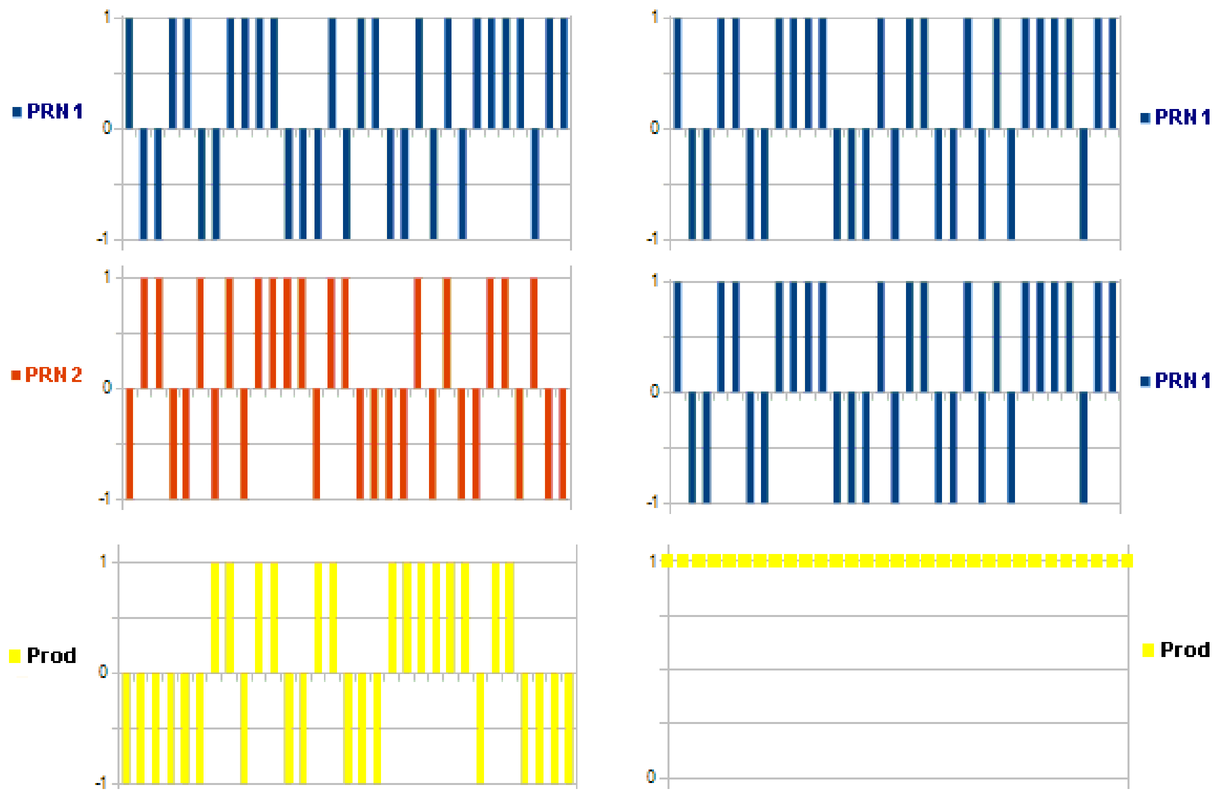


Since the GPS receivers “know” the original PRN code, they simply reapply the PRN code to the very weak incoming signal, shifting it slightly over many iterations until the receiver's PRN code aligns with the incoming signal's PRN code. This is analogous to trying to align a picket fence with missing slats (where the missing slats represent 0's in the bit stream) with an exact copy of itself. When they are not yet aligned, the sum of the products of all solid slats (“1”s) multiplied against missing slats (“0”, or “-1” in the figures) sums roughly to ~zero. When all of the slats align, the sum of the product of all the bit products produces a large total that effectively flags the alignment as being a match. When the two sequences become aligned, the receiver has effectively sync'd up to the GPS system timing. Being sync'd up with the GPS system time allows it to now begin reading the data embedded in the 50 b/s stream.

So how does spreading our data stream boost its signal strength? It doesn't. In reality, it is boosted *relative to the noise* (i.e. it "pushes down the noise"). In the parlance of radio communications, spreading increases its *Signal-to-Noise ratio*, "SNR".

At the transmitter the 50 b/s bit stream is convolved against the PRN code for the *first* time, spreading its energy out across the channel. At the receiver the 50 b/s stream is convolved against the PRN code for the *second* time, despreading its energy back to its narrow (and thus concentrated) 50 b/s bandwidth. Any ambient noise that gets picked up at the receiver on the channel is at the same time put through this same convolving process against the PRN code. However since the ambient noise is now being convolved for the *first* time it gets spread out (just as did the original 50 b/s signal back at the transmitter). As a result, the despread 50 b/s GPS signal is now some 20,000 times (43 dB =10 log [1.023 Mbps/50bps]) stronger than the [spread] ambient noise !





Each PRN code has a specific and well-defined pattern of 1's and 0's. In addition, these PRN codes are “noise-like”, meaning they contain *roughly* an equal number of 1's as 0's. As a result, when you multiply one PRN code against another (or against a shifted version of itself), the resulting summed bit products is roughly zero. However when you multiply a PRN code against an exact copy of itself, the sum of the bit products is huge, and now “*bang*” it *immediately* knows it has established a “lock”. Consequently, to extract the weak GPS signal from the noise, the receiver simply applies the known PRN code to the “noise” it hears on the channel and sums up each bit comparison. If no match is found, it shifts its copy of the PRN code slightly and tries again. It continues this process until it either gets a lock, or it eventually drains its battery and dies.

Mathematically, if we use “d(t)” to designate our 50 b/s data stream, PRN1 as the convolving code used at the SatV, and PRNz as any other PRN code (or a shifted version of PRN1), when we multiply and integrate these out, we get:

$$\sum d(t) \times PRN1 \times PRNz = d(t) \times 0 \approx 0$$

$$\sum d(t) \times PRN1 \times PRN1 = d(t) \times 32{,}000 \ !$$

(where for simplicity we assumed our PRN code in this example is 32,000 bits long; see Appendix A for additional mathematical details).

Selective Availability (SA) and Differential GPS (dGPS):

In an attempt to control the accuracy of the GPS system against the possibility that an enemy might attempt to use it against the US or our allies during war time, the design of the GPS system included an ability to degrade the accuracy of the C/A code transmissions. This degradation, known as "Selective Availability", can be achieved through little more than "dithering" or otherwise randomizing the system clock transmitted on the C/A code. In the early 1990s during the first Gulf War, Selective Availability was activated, causing many of the consumer grade GPS receivers to become almost useless. With a great many users now having grown dependent on GPS, this sudden loss effectively motivated the private sector to develop work-around solutions to negate the effectiveness of SA, while at the same time increase the accuracy of the GPS receivers utilizing these work-around solutions.

One of the more successful techniques developed in this effort is what is known as "Differential GPS" (dGPS). In a dGPS system, a specially equipped "base station" GPS receiver is set up over a known location and allowed to monitor the GPS signals. It then compares the normal GPS solution generated using broadcast GPS signals against its known location, generating a set of correction factors for each SatV signal based on the difference between the two location values. The base station then begins broadcasting these correction factors out to other specially equipped GPS receivers over a local VHF radio channel. The monitoring receivers in this network then use these correction factors to refine their location solutions. With this technique, these dGPS systems are able to increase the accuracy of the GPS location solution down to within a few inches or centimeters.

Recognizing the usefulness of such a highly accurate system in a variety of applications (e.g. civil engineering land surveying, etc.), several countries (Japan, US, as well as the EU) have now launched geosynchronous satellites which continuously generate and broadcast these dGPS signals. Such "Satellite Based Augmentation Systems" (SBAS) allow next generation GPS receivers to provide even more accurate location solutions when inside the footprint of these dGPS satellites, compared to the first generation GPS receivers.

FYI, if you have the time and need to generate a highly accurate location for some key point of interest, a very simple yet effective technique is to mount the GPS receiver at that location and capture *many* hours of location data from several different days (preferably during periods of low PDOP). Once you have that data, simply average out all the latitude, longitude, and altitude values (e.g. dump the data into a spreadsheet and use the average function). This technique averages out all the truly random variations in the data, leaving you with a much more accurate location value. (Note that this technique can be used to improve the accuracy of any data sample corrupted by any kind of truly random effects; it cannot however remove any consistent bias errors such as measurements taken with a warped yardstick, etc.).

The Ionosphere:

One of the most significant sources for uncertainty in the "time of flight" measurement is the previously mentioned transient fluctuations in the Ionosphere – or more precisely, in transient fluctuations in the electron concentrations in the Ionosphere.

Extensive studies of the Ionosphere have shown that the propagation delay of high frequency electromagnetic waves moving through the ionosphere is directly related to the variable density of electrons along their path:

$$T_{delay} \approx k*TECU / Freq^2_{carrier}. \qquad \text{Eqn 1}$$

(where k ~ 40.3, TEC = "Total Electron Concentration" and one TEC Unit = 10^{16} electrons/m^2). Therefore any indeterminate "transient" fluctuations in TEC, corresponds to an increase in the uncertainties related to "time of flight" propagation delays.

Long-term total Electron Concentration in the ionosphere varies under three different time scales:

1) by time of day (local noon being the daily peak),
2) by season (greatest at the March and September equinox), and
3) during peak solar flare activity (which tends to follow an eleven year cycle, 2013 being the most recent peak).

TEC also fluctuates over very *short* time frames, primarily due to what is known as "Traveling Ionospheric Disturbances" (TIDs). Research indicates that TIDs are generated by what is known as "Acoustic-Gravity Waves" (AGWs) propagating through the atmosphere (e.g. Fedorenko et al). These disturbances represent spatial quasi-periodic modulations of electron density "N" along a layer of the atmosphere concentric with the earth's surface. These fluctuations effectively vary the index of refraction "n" (where n = c/v, with "c" being the speed of light in a vacuum and "v" being the speed of the electromagnetic wave through a medium denser than a vacuum). This non-deterministic fluctuation of the index of refraction results in non-deterministic variations in the radio wave propagation times for all high-frequency radio waves crossing this region.

There are two primary sources of AGWs and thus TIDs: 1) energy / momentum transfer during high solar wind activity (i.e. solar proton impact, primarily at the Earth's poles), and 2) from intense terrestrial storms from equatorial to higher latitudes (hurricanes and tropical storms). AGWs and thus TIDs can also be caused by "anomalous" high energy, high impact events (e.g. during the testing of thermo-nuclear warheads, cometary impact, etc.).

In the literature, TIDs are typically separated into two classes, based on wavelength (spatial period) “L”: 1) Large-Scale (LS), and 2) Medium-Scale (MS) traveling ionospheric fluctuations. The LS TIDs are characterized by wavelengths L > 1000 km, temporal periods T ~ 0.5–4 h, and horizontal velocities V ~ 0.4– 1000 km/s. The corresponding parameters for MS TIDs are L ~ 100–600 km, T ~ 0.25–1 h, and V < 0.25 km/s. Researchers have predicted theoretically (Hocke & Schlegel) and proven experimentally (Fedorenko et al) that minimal wavelengths L for MS TID do not exceed 100–150 km.

Research carried out within the last six decades has revealed that MS TIDs occur much more frequently than LS TIDs. With rare exceptions, LS TIDs move from polar regions to the equator, they are observed often, but not always, after auroral disturbances during magnetic storms (Francis 1975; Hocke & Schlegel 1996). Long-term experimental studies using satellite beacons, indicate that both LS and MS Traveling Ionospheric Disturbances are of the same nature but only of differing wavelengths, and thus only represent different phases of disturbance dynamics.

Due to their large horizontal scales, LS TIDs can only propagate in the earth-ionosphere waveguide, while MS TIDs can move parallel to the earth surface in guided wave-mode (e.g., Gossard & Hooke 1975), as well as in free wave-mode (e.g., Francis 1974). In the latter case, TID waves are reflected by the Earth’s surface.

Ionospheric Scintillation[2]:

In addition to the uncertainties introduced in the “time of flight” measurements due to TID related index of refraction, random “scintillation” effects are known to produce transient variations in the GPS signal strength as well as its path delays. These Ionospheric scintillations are typically rapid temporal fluctuations in both the amplitude and the phase of satellite signals as they pass through the Ionosphere due to refracting through fluctuations/irregularities in the electron distributions encountered along their propagation path. The amount of scintillation varies dramatically from one day to the next, with the most severe scintillations being observed near the poles (at auroral latitudes) and near the equator (within ± 20 degrees of geomagnetic equator) due to energy/momentum imparted into the Ionosphere from solar winds, and from tropical storms (respectively; see previous discussion on TID).

Equatorial amplitude scintillation affects both GPS code and carrier tracking, and can thus have a degrading effect on pseudo-range measurements. Occasional deep scintillation fades can cause loss of lock in both code and carrier tracking within a GPS receiver. Equatorial phase scintillation adversely affects the operation of a receiver’s phase lock loop (PLL) and can lead to carrier cycle slips, navigation data bit errors, and complete loss of carrier lock.

2 Scintillation: “twinkling”, flashing or otherwise tremulous effect of an observable source/signal.

Conclusion / shameless plug:

This concludes our whirlwind tour of GPS, and we hope you found its information helpful. If so, please take a moment to visit our website to leave your comments (and/or suggestions for other possible projects). While there feel free to check out our other technical tutorials, books, and electronic hobby projects, as well as a few of our novels.

"Happy Trails".
Steve Lee

EpiphanyBySteveLee.com

References:

EpiphanyBySteveLee.com (for additional technical tutorials, etc.).

Steve Lee, *Spread Spectrum CDMA*, McGraw-Hill 2002. (For a more thorough description of Spreading Gain, CDMA, and wireless digital communications in general).

www.Trimble.com (for additional information on GPS, PDOP predictions, Ionosphere TEC modeling, etc.).

Y.P. Fedorenko et al., *Model of traveling ionospheric disturbances*, Published by EDP Sciences 2013; J. Space Weather Space Clim. 3 (2013) A30, DOI: 10.1051/swsc/2013052.

Fedorenko, Yu.P., O.F. Tyrnov, and V.N. Fedorenko, Parametres of empirical model of traveling ionospheric disturbances, Elektromagnitnye volny i elektronnye sistemy, 13, 21–46, 2008.

Fedorenko, Yu.P., O.F. Tyrnov, and V.N. Fedorenko, Parameters of traveling ionospheric disturbances estimated from satellite beacon observations in low Earth orbit, Geomagn. Aeron., 50, 489–503, 2010a.

Hocke, K., and K. Schlegel *A review of atmospheric gravity waves and traveling ionospheric disturbances*: 1982–1995, Ann. Geophys.,14, 917–940, 1996.

Appendix A: Miscellaneous *stuff* (and *things*):

"ρ" or "rho" – psuedo range (distance measured to each SatV), which depends on two things:

DOP - Dilution of Precision (no units; depends on SatV constellation geometry)

URE -User Range Error ("signals in space" error sources, e.g. ionosphere, ephemeris age, clock err, etc.)

UEE - User Equip Error (e.g. rcvr noise, multipath, EMOI/RFI, etc.)

HDOP = SQRT($northDOP^2 + eastDOP^2$)
PDOP = SQRT ($HDOP^2 + VDOP^2$) = SQRT($\sigma^2_x + \sigma^2_y + \sigma^2_z$)
TDOP = SQRT(σ^2_t)
GDOP = SQRT ($PDOP^2 + TDOP^2$)

Total error => URE * GDOP = URE * sqrt($PDOP^2 + TDOP^2$)

dGPS - Differential GPS (augmented GPS using local corrections)
SBAS - Satellite Based Augmentation System (e.g. WAAS, QZSS)
WAAS - Wide Area Augmentation System
QZSS - Quasi Zenith SatV Sys (proposed 3 SatVs)

Mathematical representation of the GPS signal:

$$S_1^p(t) = A_p P^p(t) D^P(t) \cos(2\pi f_1 t) + A_c G^P(t) D^P(t) \sin(2\pi f_1 t)$$
$$S_2^p(t) = B_p P^p(t) D^P(t) \cos(2\pi f_2 t)$$

where:

A_p, A_c = amplitudes (power) of P(Y) - code and C/A - code

$P^P(t)$ = pseudorandom P(Y) - code

$G^P(t)$ = C/A - code (Gold code)

$D^P(t)$ = navigation data stream