GPS Explained

Last Updated Monday, 02 June 2008

Introduction Satellite navigation systems are systems that use constellations of satellites each transmitting a signal to enable a user with a receiver to accurately position themselves on the surface of the earth. The American GPS (Global Positioning System) will be the focus of this article. Simple Explanation The hand held GPS receiver calculates its position by first measuring the distance between itself and 4 satellites within the Satellite Navigation constellation (Four satellites are required as there are 4 unknowns that need to be calculated, the x, y, z, and time). This is done by measuring the time delay between the transmission and reception of each satellites signals, as the speed at which the signals travel is known (near the speed of light).

In addition the satellites also transmit system health information as part of the navigation message, rough information on the orbits etc of all the satellites in the system known as the almanac and precise information on the satellites orbit, known as the Ephemeris (the satellite only transmits its own Ephemeris but all satellites transmit the almanac which contains system wide information) .

The next step is to convert the calculated position into a local reference frame, this is done via a geode specific transformation.

Detailed Explanations The GPS satellites transmit two signals one is freely available, known as the C/A (Coarse/Acquisition) Code and is used for the Standard Positioning Service (SPS), and gives an accuracy of 5 - 10m. The second is the P (Precise) Code, used for the Precise Positioning Service (PPS) , this is reserved for US military use and is encrypted to prevent spoofing/unauthorized usage, this can give an accuracy of a few meters. Each satellites C/A code is different (generated via a Pseudo Random Number generator) so that each satellite can be uniquely identified and its information distinguished from the rest.

The P-code is decrypted by use of a Y code to produce a P(Y) code. This P(Y) code can only be decrypted by by units with access to a valid decryption key.

GPS satellites broadcast there signals on two carrier waves, known as L1 and L2;

- L1 (1575.42MHz) – Navigation Message, C/A Code and Encrypted P(Y) code
- L2 (12227.60 MHz) – P(Y) Code

Dilution of Precision (DOP) The DOP is considered to be an indication of how good the geometry of the currently visible satellite constellation is for calculating a position solution. The values that are used are indicators of how suitable the current geometry is for a position solution. A high DOP is bad whilst a low DOP is considered good. The DOP is considered to magnify the errors that are inherent within a position solution. The DOP can also be affected by structures that obstruct visibility to the sky. There are several different types of DOP, each used to measure the quality of the satellite constellation for different purposes:

- GDOP (Geometric DOP) – applies to the complete 3-D solution and time.
- PDOP (Position DOP) – applies to the 3-D position solution.
- HDOP (Horizontal DOP) – Applies to the 2-D (X & Y) position solution
- VDOP (Vertical DOP) – Applies to the height of the solution
- TDOP (Time DOP) – Applies to the timing component of the solution

As a general rule of thumb a DOP of between 1-6 are considered good whilst values greater than 6 are considered suspect and should not be used. Positioning Methods Stand Alone This is the basic method used for GPS positioning and makes use of code-generated pseudo-ranges. It is usually used with a single, non-carrier phase GPS receiver. This method of positioning can only obtain a position accuracy of approximately 5 – 10 meters for civilian usage, however greater accuracy can be obtained with access to the military codes. The receiver picks a series of satellites to listen to based on the downloaded almanac information, as each satellite in view is detected it is identified by the satellites distinct C/A code and internally an identical code it produced using the same seed number used by the satellites signal generator. These two signals can then be compared to provide a delay between the received signal and the locally generated signal. The resulting range value is known as a Pseudo-range.

Pseudo-ranges, and can be thought of as a range value that does not take into account the clock offset between the receiver and the satellite, and as a result it can never be entirely accurate. This is because the clock that the GPS receivers use is typically a cheap crystal clock that is only approximately synchronised to GPS time. It is by taking several such measurements (four at the minimum, to solve the four unknowns; three unknowns for position and one unknown for time) that the timing errors can be removed. This is easily done, as the receiver clock offsets are the same for every satellite as the signal from each satellite is received simultaneously.

It is this type of positioning that we are focusing on in this document. DGPS Initially the GPS positioning signal available to civilian users was deliberately degraded. This capability was known as Selective Availability (SA), and was intended to prevent enemy forces using GPS for accurate positioning. With SA active the best positioning solution available with a stand-alone receiver was approximately 100m. SA was eventually turned off during the year 2000, however is did spur the development of a powerful tool for improving upon GPS position solutions.   DGPS is based on the use of multiple receivers (two or more). One receiver is kept stationary at a known point, and acts as a reference, whilst the rest can move around. As the position of this reference point is known corrections to the position solution that it generates can be calculated. These corrections can then be transmitted to the roving stations via a radio link so that they can improve their position solution. This technique is only really accurate up to about 500km differences between the roving stations and the reference stations, as beyond that atmospheric conditions and the satellites in use are increasingly likely to change. Carrier Phase   Carrier phase (some times known as relative) positioning is another method of using the GPS signals to obtain a position solution. It does not rely on the pseudo-random number code generated by the GPS satellites to produce it solution, it relies on the phase of incoming carrier signals at two or more receivers. It is the method of positioning most widely used for surveying applications and is accurate to a few millimetres.

The phase angle of the incoming wave is measured, this information can be combined with knowledge of the wavelength of the carrier wave to produce a position solution. The distance between the receiver and the satellite is a number of waves (known as the “integer ambiguity”) plus the fraction of the wave that is received at the receivers end (sometimes known as the “carrier phase”).

This method of ranging is significantly more difficult to perform than simply using the code pseudo ranges, not only in terms of processing the data but also as a continuous “lock” has to be maintained with the various satellites that are being used for the positioning solution. A loss of lock (known as a “cycle-slip”) means that the process of obtaining the integer ambiguities has to be started again from scratch as for all intents and purposes there is a new integer ambiguity value.

It is now possible to perform this type of positioning in real-time, which has lead to the development of the Real-time Kinematic positioning technique. These improved systems are possible due to technological advances allowing the real-time transfer of data over short distances that are essential to some of these systems, increases in the available computing power within the receivers themselves due to advances in micro-processor design, and improvements in the mathematical techniques that are used to obtain the positioning solutions. Augmentation Systems Also known as overlay services they come in two varieties, satellite based (SBAS) and ground based (GBAS). They are required because satellite navigation systems are not designed with civilian applications in mind.

Civilian applications require improved performance over that currently available, guarantees that the system performance is within the operational parameters of the system and the system is not being degraded in some way and they also provide someone to blame if things go wrong (liability).

To provide solutions to these requirements augmentation systems usually provide DGPS corrections along with information on the reliability of a positioning solution. In addition to this the space based augmentation systems are able to provide an additional ranging source from there own satellite by broadcasting a GPS like signal, thus increasing the availability of a positioning signal. Sources of Error GPS is a powerful tool for positioning, but it is also subject to several systematic errors and random noise that must be dealt with in order to obtain an accurate positioning solution.

These error sources can be split into three broad groups, known as;

- Satellite Errors – Clock & Ephemeris errors
- Propagation Errors – Tropospheric & Ionospheric refraction
- Receiver Errors – Clock bias, Antenna phase centre variations, calculation errors & Multipath.

Most of these errors have already been taken into account by a GPS receiver and there effects reduced as much as possible. Over all the errors (even when compensated for) add up to between 10-15 meters of error in a given stand alone solution. These can be improved upon by the use of augmentation services to approximately 4 meters. Satellite Clock Errors Whilst the atomic clocks on GPS satellites are incredibly accurate they do suffer from some error (mostly caused by relativistic effects). This must be taken account of in the calculations for producing a GPS position solution. (Clock errors introduce approximately ±2m of error into a given solution). It should also be noted that the effect also causes some errors but these are not normally calculated due to the complexity of the calculations required and the very small effect these have on the final position solution. Ephemeris Errors The Ephemeris information transmitted by a satellite contains some errors. This is due to orbital drift, and attitude corrections by the satellite itself. The ephemeris information is periodically updated from ground stations with information obtained by laser/radar ranging techniques to find the precise satellite location. (Ephemeris errors introduce approximately ±2m of error into a given solution). Tropospheric Errors The troposphere is generally considered the lower part of the earth's atmosphere. The main attribute that affects GPS signals here is Humidity (but temperature and pressure also have an effect) which caused the signal to be refracted. In order to correct for this detailed meteorological information is required and a mathematical model for defining how these physical parameters affect the signal.(Tropospheric effects introduce approximately ±0.5m of error into a given solution). Ionospheric Error The Ionosphere is essentially a layer of charged particles above the troposphere. These cause the speed of the GPS signal to slow down and the signal to be refracted, thus causing errors in the range values obtained. This can be compensated for to some extent by using both the available frequencies and comparing there arrival time (because the lower frequencies are affected to a greater degree by this effect). Additionally models of the Ionosphere and detailed information can be used to reduce this source of error if unable to receive the 2 frequencies.(Tropospheric effects introduce approximately ±5m of error into a given solution). Clock Bias Placing an atomic clock within a GPS receiver is not feasible. Therefore the GPS receiver has to rely on a much more inaccurate clock, this can introduce errors into the final positioning solution obtained. Antenna Phase Center In order to obtain millimetre accuracy from a GPS solution requires the antennas to be used to have their exact electrical centre known This is considered to be point for which we are computing a position solution. This can change due to satellite position and can vary between different types of antenna. Multi-path This is caused by the GPS Signal being reflected of surfaces such as tall buildings and rock faces. This will increase the travel time of the signal and thus cause errors. These can be reduced by the use of mathematical models and special antenna designs (Multipath can introduce approximately ±1m of error into a given solution). Calculation Errors Finally when performing the calculations and the rounding numbers within the calculations can cause and additional ±1 meter of error to be introduced onto the final position solution. Indoor GPS Solutions The use of GPS systems indoors is problematic as the system was designed primarily for out door use. Once indoors the signals that are received are far weaker, and they will suffer from excessive multipath effects (in many cases multipath signals are required in order to get enough satellite signals for a position solution)

To solve these issues requires the use of A-GPS and the ability to receive very weak signals GPS signals. Assisted GPS (A-GPS) An A-GPS receiver is able to download the information it would normally have to download from the GPS satellite constellation via a much faster wireless data link (the ephemeris information can take several minutes to download from the satellite signals, whilst via A-GPS it can take a few seconds). This allows the acquisition times for the satellite signals to be significantly reduced and provides enough information for the GPS receiver to enter its signal tracking mode much sooner. Weak GPS Signals In order to effectively receive and process the weak GPS signals requires the use of large banks of correlators (these are the devices that measure the difference between an incoming signal and a stored reference signal, essentially these are the bits in a GPS receiver that detect the signals). This process is generally known as Massively Parallel Correlation.

This allows the dwell times during the search for a satellite signal to be significantly increased, thus increasing the receivers sensitivity (with enough correlators all possible options can be calculated and searched nearly simultaneously).

It is by this process that many copies of the GPS signal are acquired every second, these can be used to construct a complete copy of the code that a satellite is transmitting. Putting It Together.. It is by the use of these two technologies that a comprehensive indoor GPS solution can be obtained, and whilst they can be used independently it is much more powerful when the two technologies are used in complementary fashion. The A-GPS is used to provide corrections, ephemeris and almanac information required for the position solution, whilst the high sensitivity GPS receiver is used to rapidly construct the satellite signals from which the GPS solution is calculated.