Where on Earth am I? Don't Worry, GPS Satellites will Guide you

2. Mechanism and Uses of GPS

Makarand Phatak

With the help of a couple of satellites going around the earth we can find our position on the earth with very high accuracy by measuring ranges from these satellites. The satellite based navigation system is called the GPS (global positioning system). In the first part[1] we described the basic principle of the operation of GPS. This part is about the mechanism and uses of GPS.

Makarand Phatak is working as a researcher at Aerospace Systems Private Limited, Bangalore. He received his PhD from the Indian Institute of Science in 1992. Impressed by the statement "The best way to learn is to teach and the best way to teach is to learn" he seeks joy in teaching and learning within his capacity.

In the first part of the article we have seen that to determine two components of position (latitude and altitude) in a two dimensional space, a minimum of three satellites is required. An extra satellite is required because one clock offset also comes into the picture along with two components of position. But the space is three dimensional and position has a third component (longitude). To determine the three components of position and one clock offset, a minimum of four satellites is required.

The general problem of locating the antenna of the receiver in three dimensions from the given pseudo-ranges and locations of four satellites is geometrically illustrated in *Figure 1*. The coordinate frame shown is the reference frame used by GPS, it is called earth centered earth fixed (ECEF) frame. As the name suggests the origin of the frame is the centre of the earth. Also, it is fixed to the earth, which means that as the earth rotates the reference frame also rotates with it. Apart from the reference frame for the three dimensions of space GPS has a reference time, called GPS time, maintained on the earth at Colorado Springs, USA. The satellite clocks, though very accurate, are different pieces of equipment and so are not, in general, exactly synchronised with each other. Thus, they have different offsets with respect to GPS time. A basic measurement in GPS is that of

---

[1] Part 1 of this article appeared in *Resonance*, Vol.3, No.8, 1998.

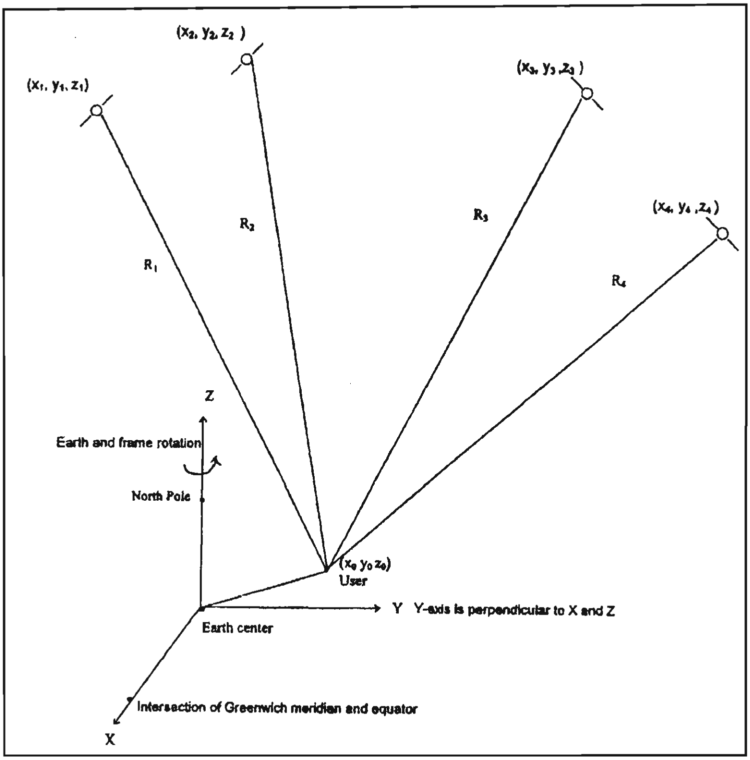


Figure 1. Positioning in three dimensions in ECEF frame.

'transmit time' to a satellite. Pseudo-range is obtained by subtracting the transmit time from the 'receive time' and multiplying the result by the speed of the signal (which is speed of light). Thus, $\rho_i = c\,(t_0 - t_i)$, where ρ_i is pseudo-range, c is speed of light, t_i is transmit time and t_0 is receive time. Further, $t_i = t_i{*} + \delta t_i$ and $t_0 = t_0{*} + \delta t_0$, where $t_i{*}$ and $t_0{*}$ are transmit and receive times respectively with respect to the GPS reference clock (at Colorado Springs) and δt_i and δt_0 are the offsets of clocks on the satellite and at the receiver respectively with respect to the GPS reference clock. Then pseudo-range is $\rho_i = c\,(t_0{*} - t_i{*}) - c\,\delta t_i + c\,\delta t_0$. The term $c\,(t_0{*} - t_i{*})$ is nothing but the range R_i. Using the formula for range R_i the basic equation of GPS becomes

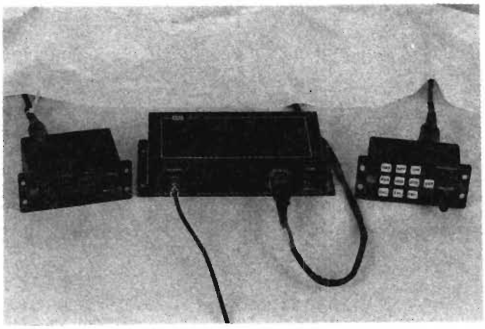
$$\rho_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2} - c\delta t_i + c\delta t_0$$

$$i = 1, \quad ., n, n \geq 4.$$

In the above equation ρ_i is obtained from $\rho_i = c\,(t_0 - t_i)$, the satellite positions (x_i, y_i, z_i) and clock offsets δt_i at transmit

(P_1) GPS receiver for military aircraft (display, receiver and keyboard units are at different locations inside a cockpit).

P_1



times can be found out (the former by solving Kepler's equations and the latter by evaluating clock polynomials), and the position of the receiver's antenna (x_0, y_0, z_0) and receiver's clock offset δt_0 at the receive time are the unknowns. The transmit times, parameters of the Kepler orbits and clock polynomials are obtained from the GPS signal. What is the nature of the GPS signal? Who determines the parameters of the Kepler orbits and clock polynomials? Now we address ourselves to these questions.

Mechanism of GPS

We have seen that to determine three components of position and one clock offset a minimum of four satellites is required. A constellation of 24 GPS satellites has been set up in space by DoD, USA, which guarantees a minimum of four satellites to be visible above the horizon anytime and anywhere on the earth. The GPS satellites are not stationary relative to the earth. Each GPS satellite completes two revolutions round the earth when the earth completes one rotation. Before one or more GPS satellites set below the horizon and become invisible at some place, other GPS satellites would rise above the horizon and the minimum visibility of four satellites is maintained. GPS satellites are arranged in six orbital planes; *Figure 2* shows GPS satellite constellation with 24 satellites in 6 orbital planes. The orbits of the GPS satellites are at an average height of about 20000 Kms. They are nearly circular and have an inclination of nominally 55° relative to the equator.

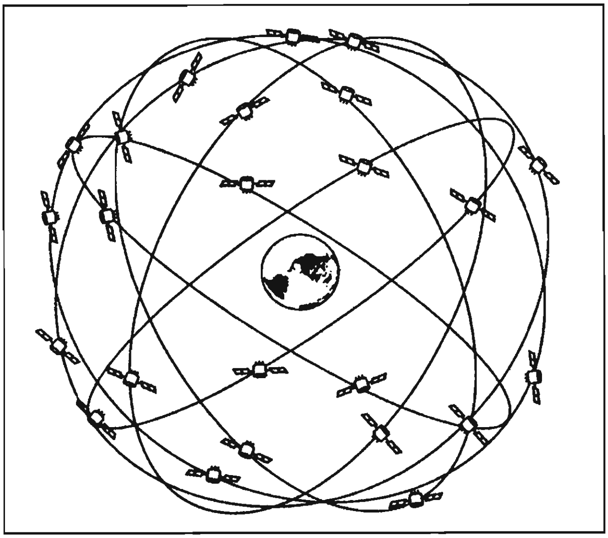


Figure 2. GPS satellite constellation.

When a GPS satellite moves along its orbit its position continuously changes with time relative to the earth, and to determine its position at a given time we need precise information about its orbit. Further, the clocks of the satellites, though very accurate, have small offsets with respect to the GPS time and these also must be determined. To find orbital parameters and the satellite clock offsets five monitoring stations are spread over the earth at known locations. These stations continuously receive signals from GPS satellites, calculate pseudo-ranges from the satellites, and relay this data to the master control station in Colorado Springs, USA. The master control station maintains the reference GPS time using the master set of atomic clocks, processes this data to determine the parameters of the predicted orbit and the predicted satellite clock offsets (typically 4 to 6 hours in future)and uplinks this information to the GPS satellites. This information is then transmitted by the satellites as part of the navigation message of the GPS signal.

The GPS signal has a carrier frequency of 1575.42 MHz, called L1 (there is one more carrier frequency of 1227.6 MHz, called L2; more about it later). L stands for L band of frequencies extending from 1000 MHz to 2000 MHz. The signal has two types of binary bi-phase modulations: code and data. Each satellite uses a distinct code. Code is a binary sequence of two states (say zero and one). The navigation message is nothing but the data and it is also represented as a binary sequence of two states. The frequency of state transition is approximately 1 MHz for code and 50 Hz for data. Multiplication of code and data produces a stream of ones and zeros and whenever in this stream there is a transition from zero to one or one to zero, the phase of the sinusoidal carrier signal changes by 180°. Distinct codes make signals distinguishable by their detailed coded structures and allow all the satellites to transmit at the same frequency and time. This type of transmission is called CDMA (code division multiple access). It does not affect and is not affected by other transmissions in the same frequency band at the same time. The transmitted signal travels nearly 20,000 Kms before reaching the receiver. As a result of propagation over such a large distance attenuation of signal power takes place and in fact the power in the signal at the receiver is less than the power of the receiver noise produced by random electronic phenomena. Then to receive signals of enough energy, is a large antenna required? The answer is no. The code based signal processing in the receiver can recreate a replica of the signal embedded in the noise. In fact, CDMA transmission enables a receiver to operate under very noisy environments. Every GPS satellite actually uses two code modulations: C/A (coarse acquisition) and P (precise) code. The frequency of state transition of P code is ten times the corresponding frequency of C/A code. The C/A code repeats every one millisecond and the P code repeats after a very long time of about one week. Thus the length of a binary sequence of C/A code is considerably smaller than that of P code. So, initially it is much easier to search presence of a particular C/A code in the incoming signal and then handover to P code. As the name suggests P code is expected to give precise range information but it is a secret. Thus precise service from GPS is not open to all. (Actually P code is no more a secret, more about it later.) The signal processing for acquisition involves the receiver generating a local replica of the code, doing multiplication of the replica with the incoming signal state by state and accumulating the results. The properties of the codes are such that only when the code patterns of replica and the corresponding component in the incoming signal match state by state (zero matches with zero and one with one) the accumulated value builds up indicating code acquisition. After acquisition, the code is tracked, i. e., the match between the incoming code and local replica is maintained in the presence of disturbances such as relative motion between the satellite and the receiver. The pattern match is done in parallel with four or more codes. The process of code pattern matching rejects most of the noise, permits collection of data from the navigation messages, and makes replicas of the clocks of the satellites available to the receiver. Note that these clock replicas show what the satellite clocks would have shown not at the present time (that is receive time) but at times in the past (that is at the respective transmit times). So, whenever the receiver needs to know the transmit times of satellites, it simply samples the locally generated replicas of the code clocks; it also samples its own clock to get the receive time. It then subtracts transmit times from the receive time and multiplies the results by speed of light in vacuum to find pseudo-ranges from the satellites. These are pseudo-ranges, i. e., ranges with offsets, because the transmitter clocks and receiver clock are not synchronised to the GPS reference time. The offset introduced by the receiver clock is an extra unknown which is to be found along with the position of the antenna phase centre of the receiver. To find a solution the positions of the satellites and the offsets introduced by the satellite clocks must also be found. The receiver gets parameters of the Kepler orbits and clock polynomials from the navigation messages of the GPS signals and using them it calculates the positions and clock offsets of the required satellites at the transmit times by solving Kepler's equations and evaluating clock polynomials.

It will be interesting to know that the satellite clocks are affected by relativity and the satellite to receiver range calculated by the receiver will have increasing error if the effect of Einstein's

Due to special relativity the satellite clock appears to be ticking at a slower rate whereas due to general relativity it appears to be ticking at a faster rate than an earth-bound clock.

theories of special and general relativity, on the satellite clocks is not considered. Due to special relativity the satellite clock appears to be ticking at a slower rate whereas due to general relativity it appears to be ticking at a faster rate than an earth-bound clock. More on this can be found in *Box 1*. The net effect is that the satellite's clock ticks about 27 nanoseconds (1 nanosecond is 1000000000-th part of a second) more every minute. As a result the satellite to receiver range as calculated by the receiver is less by 8 metres every minute or by 480 metres per hour. To reduce this increasing error due to relativity in range calculations, the ticking rate of the satellite's clock is reduced by an appropriate amount before the launch of the satellite. Still some error is left which can be compensated for approximately by using the clock parameters contained in the navigation message of the GPS signal.

Error due to relativity or any other error can be reduced but not nullified. What are the other errors? There are many. The code pattern match is not perfect, i.e., the received code and the locally generated code have some time skew. Since we have access to the local replica of the code (the received code is buried in noise) the transmit time obtained from it has some error. The position of the satellite as calculated from Kepler's orbit parameters has some error, i.e., in space the actual position of the satellite is different from the calculated position. The error is due to inability of the master control station to perfectly predict the orbit. Similarly the offset of the satellite clock with respect to the GPS time is not very precise. The satellite codes slow down while propagating through the ionosphere and troposphere. There are empirical models which can be used to estimate the delays. The ionospheric delay is inversely proportional to the square of the frequency of the carrier. It can be accurately estimated by one more measurement L2 of the range at a second frequency. Assuming other factors are approximately the same, the difference of the two range measurements (L1 and L2) is a function of the proportionality constant (unknown) and the two frequencies (known). This permits estimation of the unknown

Box 1. Einstein and GPS.

Einstein wrote while publishing his completed account of general relativity that "hardly anyone who has truly understood this theory will be able to resist being captivated by its magic". Those who care only for direct utility, not just for beauty would ask ,"What's the use?". GPS is the first engineering system to use general relativity.

Special theory of relativity leads to time dilation. Suppose that a satellite with a clock is moving approximately linearly at speed v with respect to an observer on the earth. Then elementary text books show that the clock ticks more slowly.

Now consider general relativity. Again standard text books show that the ticking frequencies of two clocks, in a gravitational field with a potential difference of $\Delta\Phi$, differ by $-\Delta\Phi/c^2$. Applying this to a clock in a GPS satellite, which is at a lower gravitational potential and a clock on the earth, which is at a higher gravitational potential, the satellite clock appears to tick faster than the earth-bound clock. This effect is opposite to that arising because of special relativity and is about twice as large in magnitude.

constant and from it the ionospheric delays. Again the estimation cannot be perfect and some error will remain. All these errors get added in the pseudo-range measurement. The errors in the pseudo-range measurements are amplified by a factor of GDOP (geometric dilution of precision, see *Box 1* of Part 1) to give a larger error in the position calculated by the receiver. With C/A code the error in position was observed typically to be in the range 15–25 metres. Though on the large surface of the earth this error is negligible yet, the DoD, USA was concerned about this low error! This fantastic accuracy can be used by the enemy for destructive work or to strengthen its defence. Initially the DoD had estimated that with C/A code the position error would be about 100 metres – a level of error which other local positioning equipment can achieve. To increase position error from 15–25 metres to 100 metres DoD then deliberately introduced additional error mainly in the timing of the signal (this error addition is known as selective availability) so that the resulting position error will increase to about 100 metres. With P code and a special electronic chip the deliberate error in the timing of the signal can be decoded and hence can be removed; with C/A code it cannot be removed. However the special electronic chip for better positioning accuracy is available only to authorised military users of USA and its allies. In fact the details of P code are now well known and documented and P code is no more a secret; however without the special electronic chip the errors in the timing of the signal cannot be decoded. With the P code known, there is a possibility that someone may transmit GPS like signals with wrong data so that the receivers are fooled and forced to give wrong positions. This is called spoofing. To prevent spoofing DoD may change P code to a secret code, called Y code. Y code signals can be received only by the receivers with the special electronic chip.

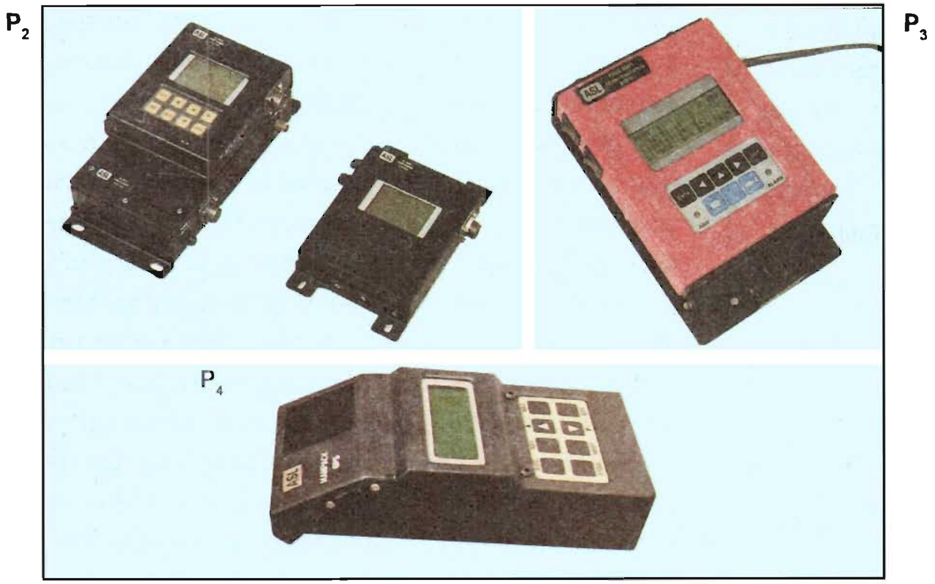
With the P code known, there is a possibility that someone may transmit GPS like signals with wrong data so that the receivers are fooled and forced to give wrong positions. This is called spoofing.

Uses of GPS

Uses of this system for military attack and defence are many. During the 1990–91 Gulf war there was an opportunity to test the efficacy of GPS by the US and its allies. GPS was successfully used in moving troops from one secret place to another during night in the unknown territory, helping other equipment to quickly locate the position of the target, destroying the early warning equipment with missiles, accurate bombing of the target despite poor visibility due to smoke, moving soldiers on planned paths in mine-embedded areas, etc.

Of course GPS has many other non-military applications. GPS is very useful in navigation in air, on sea and on land. On long distance journeys the planned path can be traversed without much deviation, resulting in saving of several kilometres and hence saving of fuel. On short distance missions also, GPS provides effective aid. For example, it helps fishermen to reach various fishing zones, the coordinates whose centres are updated roughly every week and made available at fishery offices. Photographs P1 to P4 show some of the GPS receivers built by Aerospace Systems Pvt. Ltd., Bangalore.

Though DoD has intentionally increased error globally, it is possible to reduce error locally. For this, error in range measurement is found at a location whose position is accurately known and this error, which is common in the local neighbourhood, is



(P_2) ***GPS receiver for armoured vehicles (on the right is auxiliary display).*** (P_3) ***GPS receiver for fishing trawlers.*** (P_4) ***GPS receiver for hand-held applications.***

then transmitted upto about 100 Kilometres of surrounding area. The error which still remains is only about 5 metres. This mode of operation is called differential GPS (DGPS). It opens up the use of GPS in many other applications. It is possible to have a relatively cheap yet accurate aircraft landing system. Oil exploration ships can locate an oil well, record its position accurately using DGPS and then heavy well-digging machinery can be taken to the recorded position. Using DGPS along with a map database a car's navigational computer can plan a route to the destination and guide the driver accordingly. A reference station can monitor positions of ships or trucks using DGPS and for a particular transportation task to be maximally profitable, an appropriate vehicle can be selected and a transport schedule for it can be decided. A big farmland can be partitioned and the quantity of fertiliser required for each of the partitions determined and using guidance from DGPS, the required quantity of fertiliser can be delivered to different partitions; this scheme not only saves fertiliser and money but also increases yield of the crop.

GPS can also be used as an accurate yet inexpensive time keeping equipment.

GPS can also be used as an accurate yet inexpensive time keeping equipment. The time accuracy is derived from the accurate time signals generated by satellites. GPS is becoming very popular for accurate surveying. Two receivers are used and the vector joining the antennas of the two receivers, called baseline, is very accurately obtained. If position of one end of the baseline is already known then that of the other end is obtained by vector addition. The resulting accuracy is very high; the error is as small as a few centimetres. This is achieved by exploiting the characteristics of the carrier signal rather than the code and collecting GPS measurements over a long time interval of several minutes. The baseline can be so long that the two ends of the baseline need no longer have line-of-sight connection as required in standard surveying techniques; the required connection is established through GPS satellites. It is possible to use a single receiver with multiple antennas affixed to an object and accurately determine two or more non-collinear baseline vectors. Here the antenna separation is small and the baselines are short, typical baseline length being 1 to 5 metres. Two non-collinear baselines define the orientation of the object, such as an aircraft or a satellite. Thus GPS is used not only to find position of an object but also to find how the object is oriented in space. In some high dynamic applications such as fighter aircraft, GPS is not acceptable as a sole means of navigation because it is not possible to lock on to the satellite signals especially under dynamic manoeuvers and highly jerky, accelerated and speedy motions. In such applications GPS is still very useful along with another system called inertial navigation system (INS). Stand-alone INS is not accurate; its error builds up with time. But INS-GPS combination overcomes most of the disadvantages of either of the stand-alone systems. INS helps GPS also to increase the dynamic margin under which locking on to signals can be maintained. GPS calibrates INS by estimating its errors so that when GPS is not available, due to loss of lock, the error build up of INS is reduced. GPS is useful in scientific measurements. For example it is used in measurements of gravity, of the rotation of the earth and it provides a laboratory

INS-GPS combination overcomes most of the disadvantages of either of the stand-alone systems.

for experiments in relativity.

Thus GPS has been found useful in many fields other than those which were considered in its creation. The applications of GPS will be limited only by the imagination of man.

The applications of GPS will be limited only by the imagination of man.

Concluding Remarks

Many ground based navigation systems use transmitters to transmit radio waves. The receivers of these systems receive the radio signals and make measurements which are related to the ranges to the transmitters. To solve for the position of the receiver, the positions of the transmitters should be accurately known. With the arrival of satellites it was found that the changing position of the orbiting satellite can be accurately predicted based on the laws of physics. And then it was realised that instead of stationing transmitters on the ground they can be put on the satellites to get a global navigation system. Global positioning system (GPS) is such a system which is popular and successful, using which one can obtain accurate position information anytime and anywhere on the earth.

Suggested Reading

[1] **R J Milliken and C J Zoller. Principle of operation of NAVSTAR and system characteristics.** ***Navigation. Journal of the Institute of Navigation.*** **25. No. 2, 1978.**

[2] **B W Parkinson, T Stansell, R Beard, K Gromov. A history of Satellite Navigation,** ***Navigation. Journal of the Institute of Navigation.*** **42. No. 1, 1995.**

[3] **Parkinson B W, Spilker J J (ed.),** ***Global Positioning System: Theory and Applications.*** **published by American Institute of Aeronautics and Astronautics, Inc. Vol.I and II, 1996.**

Address for Correspondence
Makarand Phatak
Aerospace Systems Private Ltd.
Pragathi, 70/1, Miller Road
Bangalore 560 052, India.

All science is concerned with the relationship of cause and effect. Each scientific discovery increases man's ability to predict the consequences of his actions and thus his ability to control future events.

Laurence J Peter