Concept of GNSS Positioning: Ranging Measurements

GNSS signals are electromagnetic waves propagating at the speed of light. Signal frequencies in the radio spectrum between about 1.2 and 1.6 GHz (a part of the so-called L-band) have been selected for these signals since these enable measurements of adequate precision, allow for reasonably simple user equipment and do not suffer from attenuation in the atmosphere under common weather conditions. At the given frequencies, GNSS signals have a wavelength of about 19–25 cm.
Similar to early radio navigation systems such as Transit, GNSSs provide signals on at least two different frequencies for compensation of ionosphere delays in their measurements.

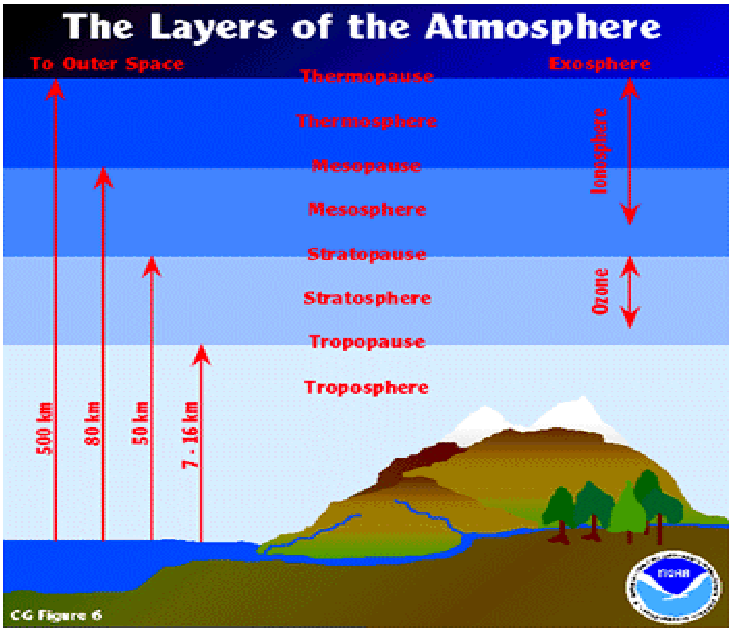

Figure (1) the layers of the atmosphere

A distinct feature of all GNSS signals is the modulation of the harmonic radio wave (termed the carrier) with a characteristic pseudorandom noise (PRN) code. This code is essentially a binary sequence of zeros and ones with no obvious pattern or regularity. The sequence is transmitted at a rate of typically 1–10MHz, where higher rates imply a higher processing effort but promise more precise measurements. The PRN code is continuously repeated at intervals of a few milliseconds to seconds and facilitates measurements of the signal transmission time. In most GNSSs, the PRN sequence also serves as a unique fingerprint, which allows the receiver to distinguish individual satellites transmitting on the same frequency. On top of the ranging code, the signal is also modulated with a low rate (e.g., 50 bits=s) navigation data stream (known as the broadcast navigation message) that provides information on the orbit of the transmitting satellite and the offset of its local clock from the GNSS system time. The basic measurement made by a GNSS receiver is the time required for the GNSS signal to propagate from a satellite to the receiver. This can be obtained by tracking the PRN code modulation of the signal as illustrated in Fig. 2. Within the receiver, a local copy of the PRN sequence is generated, which is continuously compared and aligned with the signal received from the satellite. This tracking loop provides continuous measurements of the instantaneous code phase and hence the transmission time corresponding to the currently received signal. By comparing this time with the local receiver time, the signal propagation time, and – upon multiplication by the speed of light – the distance or range from receiver to satellite are obtained. Thus, the GNSS signals enable three basic types of measurements:

1. Pseudorange:

The pseudorange is a measure of the range, or distance, between the GPS receiver and the GPS satellite (more precisely, it is the distance between the GPS receiver's antenna and the GPS satellite's antenna). As stated before, the ranges from the receiver to the satellites are needed for the position computation. Either the P-code or the C/A-code can be used for measuring the pseudorange. The procedure of the GPS range determination, or pseudoranging, can be described as follows. Let us assume for a moment that both the satellite and the receiver clocks, which control the signal generation, are perfectly synchronized with each other. When the PRN code is transmitted from the satellite, the receiver generates an exact replica of that code. After some time, equivalent to the signal travel time in space, the transmitted code will be picked up by the receiver. By comparing the transmitted code and its replica, the receiver can compute the signal travel time. Multiplying the travel time by the speed of light (299,729,458 m/s) gives the range between the satellite and the receiver. Unfortunately, the assumption that the receiver and satellite clocks are synchronized is not exactly true. In fact, the measured range is contaminated, along with other errors and biases, by the synchronization error between the satellite and receiver clocks. For this reason, this quantity is referred to as the pseudorange, not the range. GPS was designed so that the range determined by the civilian C/A-code would be less precise than that of military P-code. This is based on the fact that the resolution of the C/A-code, 300m, is 10 times lower than the P-code. Surprisingly, due to the improvements in the receiver technology, the obtained accuracy was almost the same from both codes.

2. Carrier phase:

Another way of measuring the ranges to the satellites can be obtained through the carrier phases. The range would simply be the sum of the total number of full carrier cycles plus fractional cycles at the receiver and the satellite, multiplied by the carrier wavelength. The ranges determined with the carriers are far more accurate than those obtained with the codes (the pseudoranges). This is due to the fact that the wavelength (or resolution) of the carrier phase, 19 cm in the case of L1 frequency, is much smaller than those of the codes. There is, however, one problem. The carriers are just pure sinusoidal waves, which means that all cycles look the same. Therefore, a GPS receiver has no means to differentiate one cycle from another. In other words, the receiver, when it is switched on, cannot determine the total number of the complete cycles between the satellite and the receiver. It can only measure a fraction of a cycle very accurately (less than 2 mm), while the initial number of complete cycles remains unknown, or ambiguous. This is, therefore, commonly known as the initial cycle ambiguity, or the ambiguity bias. Fortunately, the receiver has the capability to keep track of the phase changes after being switched on. This means that the initial cycle ambiguity remains unchanged over time, as long as no signal loss (or cycle slips) occurs. It is clear that if the initial cycle ambiguity parameters are resolved, accurate range measurements can be obtained, which lead to accurate position determination. This high accuracy positioning can be achieved through the so-called relative positioning techniques, either in real time or in the post processing mode. Unfortunately, this requires two GPS receivers simultaneously tracking the same satellites in view.

3. Doppler:

The change in the received frequency caused by the Doppler effect is a measure of the range-rate or line-of-sight velocity.

Pseudorange, carrier-phase, and Doppler observations provide the basic measurements for computing position and velocity as well as the offset of the receiver time with respect to the GNSS system time scale. They are complemented by information on the orbit and clock offsets of the individual GNSS satellites, which is transmitted as part of the broadcast navigation message and allows the receiver to compute the position and velocity of the transmitting satellite at the signal transmission time. To provide such information with adequate accuracy, the GNSS operator must be able to determine and to predict the satellite orbit ahead of time, so that it can be uploaded to the satellite for subsequent broadcasting to the users. Likewise GNSS relies on highly stable onboard clocks, whose time offsets can be accurately predicted. Rubidium or cesium atomic frequency standards or even hydrogen masers are used for this purpose, which deviate by only 10−13 to 10−14 from their nominal frequency over time scales of a day.

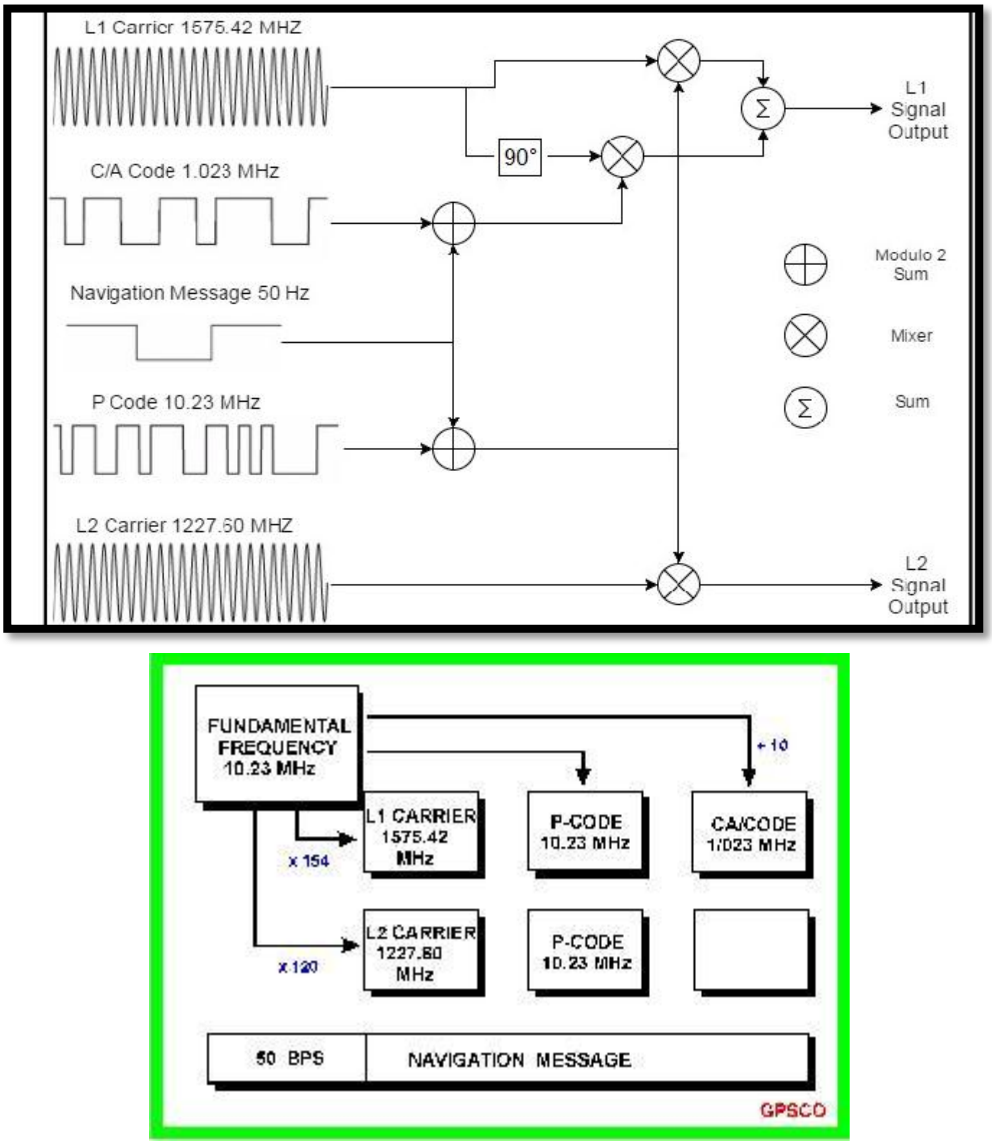

Figure (2) GNSS signal structure

Simple concept of mathematical model of the satellites range:

As previously mentioned, the signals transmitted by the satellites consist of two carrier waves (L1 and L2, in addition to other carrier waves) which are modulated with timing codes (C/A and P). Covered on these codes is a navigation message which contains information regarding the transmitting satellite's position, correction parameters for the transmitting satellite's clock and health and almanac data for all the satellites.

GNSS was designed as a ranging system from known positions of satellites in space to unknown positions on land, at sea, in air, and space. Effectively the satellite signal is constantly marked with its (own)

transmission time so that when received the signal transit period can be calculated with a synchronized receiver.
Let we present this simple operation mathematically:

$p = C.\Delta t$……………………………………………………………1

Where

p stands for the true distance between the satellite and the receiver which can be presented as:

$p = \sqrt{(x^{si} - x_r)^2 + (y^{si} - y_r)^2 + (z^{si} - z_r)^2}$…………………….....2

C represents the velocity of the electromagnetic radiation,
Δt stands for the signal travel time, which equals to:

$\Delta t = (t_r - e_{t_r}) - t^{si}$……………………………………………………….3

e_{t_r} stands for the error in the receiver clock, as the reciver
So,

$\left(x^{si} - x_r\right)^2 + \left(y^{si} - y_r\right)^2 + \left(z^{si} - z_r\right)^2 = \left(C.\Delta t - e_{t_r}\right)^2$…………..4

As it is seen that the equation number 4 contains 6 known parameters, which are:
$C, x^{si}, y^{si}, z^{si}, t^{si}$, and t_r
Where it contains only 4 unknown parameters, which are :
x_r, y_r, z_r, e_{t_r}
As a consequence, to solve this equation, 4 equations are required (at least) to calculate the 4 unknown parameters. This observation can be made either to four satellites or to one satellite in four consecutive positions ($t_1, t_2, t_3, and\ t_4$).
However, the speed of the electromagnetic radiation is significantly changed due to the change of the atmosphere. Consequently, the observation to at least four satellites represents the optimum solution to calculate the unknown four parameters, see the equation system below:

$$(x^{s1} - x_r)^2 + (y^{s1} - y_r)^2 + (z^{s1} - z_r)^2 = \left(C.((t_r - e_{t_r}) - t^{s1})\right)^2$$
$$(x^{s2} - x_r)^2 + (y^{s2} - y_r)^2 + (z^{s2} - z_r)^2 = \left(C.((t_r - e_{t_r}) - t^{s2})\right)^2$$
$$(x^{s3} - x_r)^2 + (y^{s3} - y_r)^2 + (z^{s3} - z_r)^2 = \left(C.((t_r - e_{t_r}) - t^{s3})\right)^2$$
$$(x^{s4} - x_r)^2 + (y^{s4} - y_r)^2 + (z^{s4} - z_r)^2 = \left(C.((t_r - e_{t_r}) - t^{s4})\right)^2$$

The accuracy of the user position and the receiver time can be improved significantly by increasing the number of the observed satellites which leads to increase the number of equations, where the number of the unknown parameters remain constant. In this case, the least square adjustment is required to solve the equation system and compute both of the receiver position and clock correction.
As we mentioned, the accuracy of the ranges which are measured to the GNSS satellites, are affected significantly due to lack (absence) of

synchronization between the satellite clocks and the receiver clocks. The satellite clocks are characterized by a very high stability as it based on atomic clock, where the receiver clocks are characterized by low level of stability as it depends on quartz clocks. Consequently, the range measured between the satellite and the receiver is called pseudorange as it is degraded by the absence of the synchronization between the satellite clocks and the receiver clocks.

$$\acute{p} = C.\Delta t$$