Global Positioning System

The global positioning satellite (GPS) system has revolutionized navigation and position location. It is now the primary means of navigation for most ships, aircraft, and automobiles, and is widely used in surveying and many other applications. The GPS system, originally called NAVSTAR, was developed as a military navigation system for guiding missiles, ships, and aircraft to their targets. GPS satellites transmit L-band signals that are modulated by several codes. The principles of the GPS system were made public in 1983, and GPS receivers using the C/A (*coarse acquisition*) code became available to the public by 1990. Early GPS receivers cost thousands of dollars, but prices fell quickly once volume production began. The GPS system was declared fully operational with 24 satellites in 1995. The secure high accuracy P-code allows authorized users (mainly military) to achieve positioning accuracy of 1 m. This was the accuracy that the military users wanted for targeting smart bombs and cruise missiles, but such accuracies are also useful for auto-landing aircraft in fog and for docking ships in bad weather.

The first commercial use of GPS was in surveying, but by 1995 several companies had produced low cost, handheld GPS receivers for general position location and navigation. The development of integrated circuits (ICs) specifically for GPS receivers and larger volume production quickly brought down the price of a GPS receiver, and the market expanded rapidly. GPS receivers are now a consumer product and can be found in many cars and all cellular telephones. The GPS system has been successful because it is available everywhere in the word, is free to all users, and provides a direct readout of the present position of a GPS receiver with a typical accuracy of 5m.

In 2018, the GPS space segment consisted of 27 satellites in medium earth orbit (MEO) at a nominal altitude of 20 183 km with an orbital inclination of 55°. The satellites are clustered in groups of four, called constellations, with each constellation separated by 60° in longitude, with an additional three satellites for improved coverage in certain parts of the world. The orbital period is approximately one half a sidereal day (11 hours 58 minutes) so the same satellites appear in the same position in the sky each day. The satellites carry station keeping fuel and are maintained in the required orbits by occasional station keeping maneuvers, just like geostationary earth orbit (GEO) satellites. The orbits of the 24 GPS satellites ensure that at anytime, anywhere in the world, a GPS receiver can receive signals from at least four satellites. Up to ten satellites may be visible at times, and at least four satellites are visible all of the time. Replacement satellites are launched as needed, so there may be more than 24 operational GPS satellites at any given time.

Figure 1 shows a GPS Block IIF satellite. There have been four generations of GPS satellites since the first launch of a Block II satellite in 1990. All Block II satellites were decommissioned by 2016. Block IIR, Block IIR-M, and Block IIF satellites were operational in 2018 with a total of 31 operating in space. The Block IIF satellites weighed 1408 kg at launch and had a design lifetime of 12 years. DC power of 1900W was generated by Gallium Arsenide solar cells. The satellites have three axis stabilization and multiple thrusters for orbital corrections. Fuel for the thrusters is hydrazine, with 140 kg at launch. Because GPS is an integral part of the defense of the United States, spare GPS satellites are kept in orbit and more spares are ready for immediate launch. The GPS system is operated by the US Air Force from the GPS master control station (MCS) at Schriever Air Force Base in Colorado Springs, CO, with an alternative MCS at Vandenburg Air Force Base in California.

The position of a GPS receiver is found by trilateration, which is one of the simplest and most accurate methods of locating an unknown position. In trilateration, the distance of the unknown point from three known points is measured. The intersection of the arcs corresponding to three distances defines the unknown point relative to the known points, since three measurements can be used to solve three equations to give the latitude, longitude, and elevation of the receiver. The distance between a transmitter and a receiver can be found by measuring the time it takes for a pulse of RF energy to travel between the two. The distance is calculated using the velocity of electromagnetic waves in free space, which is assumed to be equal to the velocity of light, c, with $c = 299792458\ ms^{-1}$.

It is not possible to make timing measurements with this precision with a single transmitted pulse. All GNSS satellites use direct sequence code division multiple access (CDMA) transmissions (spread spectrum) to send a long sequence of pulses in a wide bandwidth that are correlated with an identical sequence generated in the receiver. The spread spectrum signal is buried well below the receiver noise floor, typically with carrier to noise ratio (CNR) around −20 dB for the civil C/A code. The correlation process removes the code sequence from the received signal allowing it to be processed through a narrow bandwidth giving an output with signal to noise ratio (SNR) well above 0 dB and making accurate time measurements possible.

Each satellite radiates a different sequence of bits (known as chips in a spread spectrum system) which starts at a precisely known time. A GPS receiver contains a clock that is synchronized in turn to the clock on each satellite that it is receiving. The receiver measures the time delay of the arrival of the chip sequence, which is proportional to the distance between the satellite and the GPS receiver. The position of each satellite is calculated in the GPS receiver using the ephemeris for the satellite orbits that are broadcast by each satellite in a *navigation message*. Making the calculation for four satellites provides the receiver with sufficient information to determine its position with very good accuracy. Four satellites, rather than three are needed because the clock in the receiver is not inherently accurate enough. The fourth distance measurement provides information from which clock errors in the receiver can be corrected and the receiver clock synchronized to GPS time with an accuracy better than 10 ns.

Originally GPS satellites were designed to transmit two signals at different frequencies, known as L1 and L2. The L2 signal has a frequency of 1227.6GHZ and is modulated with a 10.23Mbps pseudorandom (PRN) bit sequence called the *P-code* that is used by military positioning systems. The P-code is transmitted in an encrypted form known as the *Y code,* which restricts the use of the P-code to authorized users. The L1 carrier at a frequency of 1575.42 Hz is modulated by a 1.023Mbps PRN sequence called the *C/A code* that is available for public use. C/A stands for *coarse acquisition* and P stands for *precise*. The higher bit rate of the P-code provides better measurement accuracy than the 1.023Mbps C/A code.