The Global Positioning System

Chris Williams
Student 1607421
HET 605

3rd June 2003

Introduction

Arguably one of the most useful global benefits provided by the U.S. Department of Defense (DoD) has been the Global Positioning System (GPS). The GPS provides worldwide ability to pinpoint positions, including altitude, to within a few tens of metres. The general theory of operation for GPS is relatively straightforward but the implementation must take into account a number of effects related to the special and general theories of relativity. This project describes the GPS and how relativistic effects are accommodated within it.

Theory of Operation

Traditional means of locating oneself on the Earth involved taking bearings and/or distances from known points, with the intersection of these lines/circles providing location. This method, called triangulation, works well in circumstances where the known points are visible and accurate maps are available but fails in absence of known landmarks, e.g. at sea or poor weather. The U.S. DoD required a more accurate and reliable system with global coverage and all-weather ability, and funded creation of the GPS to meet this requirement.

GPS extends the older concept of navigation into three dimensions by providing an elevated set of known reference points, satellite-borne, and means to measure their distance. Knowing the distance to a single reference point, fixes the observer's location somewhere on the surface of a sphere centred on the reference point. Adding the distance to a second reference fixes the observer's location to the circular intersection of the two reference point spheres. The addition of a third reference fixes the location of the observer to two possible points on the circle, one of which is usually unreasonable i.e. too high above ground (Figure 1). A fourth and further references may be required to resolve the location to a single point and reduce error.

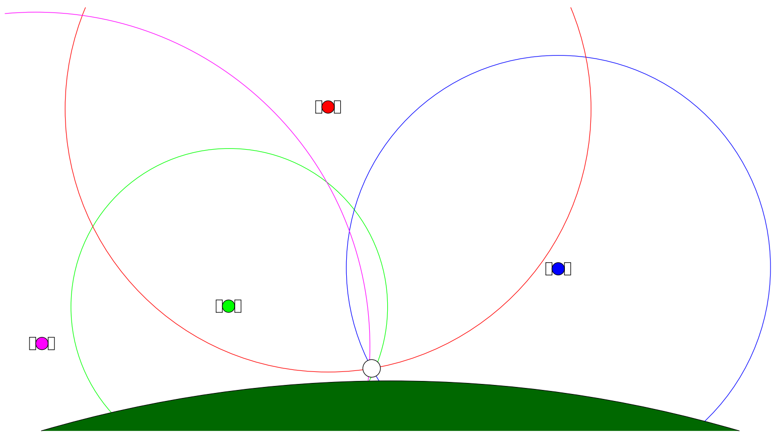

Figure 1: Ranges from three satellites coincide at two points, only one is usually sensible as a location but a fourth range can be used if needed.

The GPS reference points are a constellation of 24 satellites (21 operating, 3 spares). There are four satellites in each of six equally-spaced orbital planes inclined at 55° to the equator of Earth. The nominally circular orbit of 10,988 nautical miles (20,350 km) altitude [1] provides the satellites a 12 hour orbital period. From any point on Earth between five and eight satellites are visible at any time (closer to the poles they may be low on the horizon). The orbits also give the satellites a ground track that is close to north-south.

Distance measurement in GPS is based on one of the underlying postulates of special relativity: "The speed of light in empty space is an absolute constant of nature and is independent of the emitting body" [2]. By determining the time-of-flight of a signal sent by the reference satellite, and using the known speed of light, distance can be determined.

Light travels at $299,792,458\text{ms}^{-1}$ and will cover the satellite-Earth distance in as little as 70 milliseconds - an easily measured interval. Light, however, covers the last 100 metres in approximately 330 nanoseconds. In order to determine distances to 100 metre accuracy therefore requires determination of the time of flight to better than 330 nanoseconds. The obvious means of measuring the time-of-flight requires clocks on board the satellite and in the receiver to be synchronised to within the required tolerances. If the satellite transmits a signal derived from its time then determination of time-of-flight by the receiver becomes a case of subtracting the time of reception from the transmit time as encoded in the received signal.

All GPS distance measurement relies on synchronised clocks that remain tightly synchronised over long periods. The tight synchronisation of clocks between Earth and satellites is more complicated than it first seems. There are three issues:

- Highly accurate clocks are large, expensive, and not well suited to a handheld navigation devices. Timing on board the GPS satellites is maintained by expensive caesium beam (atomic) clocks, accurate to 2 or 3 parts in 10^{14}, approximately one second per million years. A typical receiver might use a quartz crystal-based clock sufficiently accurate only over periods of seconds.

- Identical clocks that are moving relative to each other keep time at different rates. This counter-intuitive behaviour is described by the special theory of relativity.

- Identical clocks at different places in a gravitational field keep time at different rates. This behaviour is described by elements of the general theory of relativity.

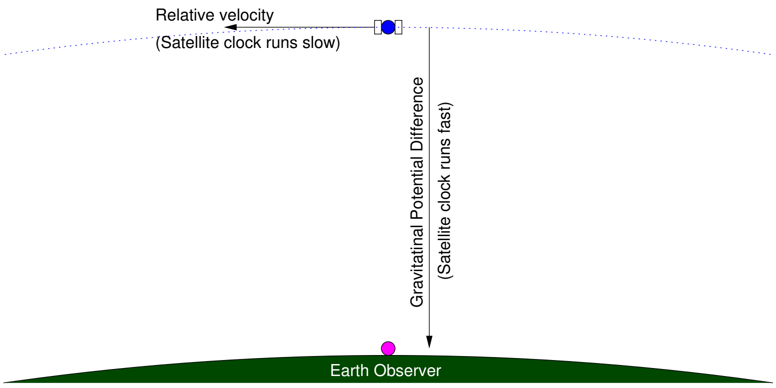

Figure 2: Two main relativistic effects. Gravitational potential and relative motion differences compete to give a net 39μs/day gain of the satellite clock over ground clocks.

The following discussion analyses the two main effects of relativity theory on the time keeping of GPS satellites (Figure 2). Time keeping will be compared to a notional clock that is inertial with respect to the centre of the Earth rather than a clock on the Earth's rotating surface. This reference frame is called the Earth-centred, Earth-fixed (ECEF) frame in GPS. Clocks on the Earth's rotating surface also differ in timekeeping from this notional clock, and this effect will be addressed later.

Classical physics tells that there are frequency changing effects as the result of relative motion. The effects are familiar as the change in pitch of siren as it passes, called the Doppler shift. In order to stay in orbit GPS satellites must maintain an orbital velocity of: $V_s = \sqrt{\mathcal{G}M/r} = 3863$ m/s, so these effects apply. Special relativity modified the classical Doppler shift equation to account for speeds that are a substantial fraction of the speed of light; the first-order Doppler shift. Einstein also identified a second-order Doppler shift induced by relativistic time dilation. The accuracy of the relativistic Doppler equation over its classic counterpart was first shown experimentally by Ives and Stillwell in 1938 [3]. The relativistic Doppler equation, expressed as fractional frequency change, has the form:

$$\frac{\Delta f}{f} = \frac{1}{\gamma(1 - \beta\cos\theta)} - 1$$

where $\beta = v/c$ and $\gamma = 1/\sqrt{1-\beta^2}$ is the Lorentz factor. The angle θ is that measured at the satellite between the velocity vector of the satellite and the observer. In the case of a notional inertial observer at the Earth's surface and circular orbits the shifts as a result of radial velocity component are symmetrical approaching versus receding, variable from place to place, and can only reasonably be accounted for by the receiver. The discussion is therefore limited to Doppler components that are fixed. When $\theta = 90^\circ$ the equation collapses to:

$$\frac{\Delta f}{f} = \frac{1}{\gamma} - 1 \tag{1}$$

which is the second-order, or transverse, Doppler shift. We therefore act as if θ is always ninety degrees and only consider the fixed second order effect. For our GPS system the second-order Doppler term is:

$$\begin{aligned}\frac{\Delta f}{f} &= \frac{1}{\gamma} - 1 \\ &= \frac{1}{1/\sqrt{1-(3863/3\times 10^8)^2}} - 1 \\ &= -8.3\times 10^{-11}\end{aligned}$$

This fractional frequency difference amounts to 7.2μs per day in time keeping. The negative sign indicates that the satellite lags behind the inertial clock.

Einstein's general theory of relativity predicts the existence of time altering effects related to the position of clocks at different places in a gravitational field. Pound, Rebka, and Snyder [4, 5] confirmed the existence of such effects over short distances in the early 1960's. Our GPS satellites are in a much weaker gravitational field than a ground or aircraft-based receiver, therefore must consider gravitational time dilation effects. A first order approximation of frequency shift induced by gravity is given by [2]:

$$\frac{\Delta f}{f} = \frac{\Phi_1 - \Phi_2}{c^2}$$

where $\Phi_1 = -\mathcal{G}M/r_1$ is the gravitational potential at the point of emission, Φ_2 at the point of reception, and positive $\Delta\Phi$ denotes red-shift. For GPS:

$\Phi_1 \simeq -14923000$ Nm/kg and $\Phi_2 \simeq -62538000$ Nm/kg (assuming sea level) and therefore the fractional frequency change is $\Delta f/f = 5.3 \times 10^{-10}$. This frequency change amounts to 45.8μs per day difference in time keeping between the satellite and Earth. The satellite clock leads the inertial clock because the number is positive.

The combination of the gravitational and relative motion effects on time keeping is:

$$\begin{aligned} \frac{\Delta f}{f} &= \left(\frac{\Delta f}{f}\right)_{grav} + \left(\frac{\Delta f}{f}\right)_{motion} \\ &= 5.3 \times 10^{-10} + -8.3 \times 10^{-11} \\ &= 4.5 \times 10^{-10} \end{aligned}$$

The satellite clock beats faster than the inertial clock by an amount equivalent to 38.6μs per day. If left unchecked, this difference would amount to a positional error of nearly 12 kilometres after just one day, which is clearly unacceptable.

One approach to countering these time altering effects is to build corrections into the algorithm used by receivers to calculate position. Alternatively, the satellite clock can be adjusted so that it beats more slowly to counteract the frequency change induced by the time altering trip to Earth. The slower rate can be adjusted so that the received clock signal beats at the same rate as a ground-based clock and therefore stays in synchronisation over longer periods. The latter option is the one adopted by the GPS; receiver corrections would be complex and increase the receiver cost, while adjusting the clock rate is relatively trivial. The GPS clock signal rates are 10.23 MHz and 1.023 MHz, although the rates are ultimately arbitrary, so the clock rate on the satellite has been adjusted to:

$$(1 - 4.5 \times 10^{-10}) \times 10.23 \text{ MHz} = 10,229,999.9954 \text{ Hz}$$

on the satellite in order to achieve desired rates at the ground.

At this point we have a theory of operation and a set of corrections applied to the clock systems of the satellite reference points in order to improve time synchronisation. The following section describes the operation of a typical receiver.

GPS Receiver

The GPS receiver has a number of tasks to achieve before a locational fix can be achieved:

- Identify the visible GPS reference satellites and determine their current location.

- Detect and synchronise with the signals from those satellites.
- Calculate the approximate distance to each satellite.
- Calculate and apply a time correction to its local clock if necessary.
- Calculate a corrected distance to each satellite and translate that to a location on or near the Earth's surface.

Each of these steps will be discussed in turn.

A basic almanac of satellite orbital information, built into the receiver, is sufficient to determine satellite visibility at any given time. The accuracy with which can be achieved is limited by the accuracy of orbital information in the face unpredictable external effects such a gravitational perturbations, solar winds etc. A series of ground monitoring stations track the satellites and calculate precise orbital characteristics which are periodically up-linked to the GPS satellites. The GPS satellites broadcast updates to the almanac to correct for any external variability and for satellite constellation changes. The receiver uses the updates, and may store them when powered down, to locate the satellites very precisely. This precise location provides the reference points necessary for triangulation.

Having identified the visible satellites, the receiver must acquire the signal from each satellite. All GPS satellites broadcast on the same carrier frequencies (1575.42 and 1227.60 MHz [6]). The receive frequency is slightly different due to the radial Doppler effects discussed earlier, allowing a degree of separation between satellites. The publicly accessible timing signal, a 1.023 MHz 1023-bit pseudo-random number (PRN) sequence lasting exactly one millisecond, is extracted from the identified carriers. The sequence is unique to each satellite and allows positive identification and lock-on to the signal.

The receiver must calculate time of flight for each acquired signal. In practise this is achieved by the receiver generating the expected PRN signal synchronised to the receiver clock and delaying it until it aligns with the incoming signal, the delay equalling the time of flight. This process is repeated for each visible satellite. The receiver now has a pseudo-range to each satellite based on the assumption that the receiver clock is correct. At least three pseudo-ranges are require to fix location and the receiver chooses the satellites that give the widest angular separations in order to maximise precision.

The assumption that the handheld receiver clock is accurate is generally a poor one. The typical quartz crystal clock would only be sufficiently accurate over periods on the order of 5 seconds [7]. Conveniently, four satellite ranges can be used to correct for any misalignment of the less accurate hand-held receiver clock. If the receiver clock is not aligned with the satellites then the fourth range measurement will not intersect with the three measurements used to determine location and the receiver can therefore recognise that it is misaligned. The receiver can then look for a common clock correction that will make all four distance measurements coincide at a single point. Having found such a correction, the GPS receiver can adjust its clock, but will continue to monitor for further drift with each position fix.

With satellite orbital data, corrected clock, and pseudo-range information the receiver can perform the necessary mathematics to locate the receiver in the ECEF frame. Manipulation is required to convert this to the Earth's surface.

Corrections and Complications

There remain a series of corrections that can be made to the basic system in order to improve accuracy. These corrections are discussed below in no particular order.

Time-of-flight calculations are affected by atmospheric and ionospheric refraction. These effects are modelled and the necessary correction factors are distributed through the satellite system to receivers using low data-rate navigation data set superimposed on the PRN signal. The navigation data include updates to the almanac and time offsets particular to the each satellite, among other things. These updates are used in refining the receiver performance.

The foregoing discussion make the assumption that GPS satellite orbits are circular and invariant. Achieving such an orbit is not possible in practise, consequently the orbits may be slightly non-circular and move over time. The introduction of eccentricity, up to $e = 0.02$, into the orbit of a satellite serves to complicate the calculation of both gravitational and transverse Doppler frequency shifts. In an eccentric orbit the altitude and velocity varies from apogee, minimum velocity, to perigee, maximum velocity. While the correction is not substantial, it is included in the almanac and associated algorithms.

Manoeuvring satellites, to swap a failed one out for example, also causes variation in timekeeping by virtue of the accelerations involved. These changes should be monitored and the affected satellites adjusted back into line with their peers before going into service.

The synchronisation of clocks in GPS satellites, and among the ground monitor stations and receivers, is vitally important. These clocks can be synchronised by light-speed signalling between them, but the Earth's rotation must be considered. Relativity theory predicts that signals travelling with the rotation, where the receiver is moving away from the signal, must travel further to reach their destination and therefore arrive later than expected on distance alone. Conversely, for signals travelling westward, where the receiver is moving toward the signal, the signals arrive earlier than expected. This is called the Sagnac effect after Georges Sagnac performed the original experiment in 1913. The effect was also demonstrated by Hafele and Keating in 1971 [8, 9] while testing time dilation using atomic clocks circumnavigating the globe in opposite directions, and is vital in the operation of laser ring gyroscopes. Ignoring the effect, which amounts to 207.6 ns [10] for a single equatorial circumnavigation, could lead to serious differences between clocks.

As a result of the Earth's rotation, ground-based clocks are also in motion with respect to the ECEF clock and therefore keep time differently. The maximum velocity a ground-based clock will have is the Earth's rotational velocity at the equator, approximately 465ms^{-1}. Equation 1 evaluates to: $\Delta f/f = -1.2 \times 10^{-12}$. This effect is two orders of magnitude smaller than that applicable to the satellites, and amounts to 104 ns or 31 m per day. Given that this is the maximum possible effect, and that time can be corrected from satellite fixes, this contribution may be ignored.

More complex analysis done by Neil Ashby [10] on the effects of relativity in the GPS indicate that other gravitational effects are worthy of consideration. The Earth is not spherically symmetrical and thus possesses a non-zero gravitational quadrupole moment. The moment describes the non-uniformity of mass distribution and is mainly due to the oblate nature of the Earth. Analysis including the Earth's quadrupole moment will not be addressed quantitatively here, except to say that its contribution does not change the scale of the gravitational effects discussed earlier.

The general theory of relativity models gravitation as a curvature of space-time. In this model the flight path of the satellite-Earth down-link signal follows a geodesic (shortest) path through space-time which may not be a straight line from the point of view of the receiver. The effect is small for Earth, but may become more significant in future, more accurate, navigation systems.

Implementations

This project has, in general, discussed global positioning using the existing US DoD GPS as the working example. The USSR launched its own global system called GLONASS because it could not rely on the US military controlled GPS. The European Space Agency intends launching the Galileo navigation system in the coming years. Galileo will inter-operate with both GLONASS and GPS systems and, crucially, not be controlled by military interests so that better accuracy is achievable by all users.

All these systems have around different design choices. However, regardless of the precise implementation, all such systems must account for the effects of relativity in their systems.

Conclusion

The implementation of a global positioning system is simple in theory but complicated in practise. Using satellites for reference points and the time of flight for light-speed signals as a measurement tactic means that several facets of relativity theory must be accounted for. The satellite clocks runs slower than their Earth-based counterparts because of relative motion. However, the slowing due to motion is countered by a rate increase induced by a weaker gravitational field at satellite altitude. The combination of these effects is faster timekeeping at the satellite, which must be counteracted if accurate synchronisation is to be maintained. The GPS does this by deliberately running the satellite clocks slower so that the received signal, at the ground, is keeping correct time. Synchronisation between satellites must be maintained and, in doing so, the Sagnac effect must be accounted for. The satellite positions and corrections for atmospheric effects must be well known to the receiver, and this is accomplished using data transmitted through the satellite system itself. Future, more accurate, systems may require incorporation of corrections for the curvature of space-time in the Earth's vicinity and the quadrupole moment of the Earth. The GPS is an excellent example of relativity at work.

References

[1] NAVSTAR global positioning system fact sheet. United States Air Force, `http://www.losangeles.af.mil/smc/pa/fact_sheets/gps_fs.htm`, May 2000.

[2] L. Sartori. *Understanding relativity: a simplified approach to Einstein's theories.* University of California Press, 1996.

[3] H. E. Ives and G. R. Stilwell. An experimental study of the rate of a moving atomic clock. *Journal of the Optical Society of America*, 28:215–226, 1938.

[4] R. V. Pound and G. A. Rebka. Apparent Weight of Photons . *Physical Review Letters*, 4:337–341, April 1960.

[5] R. V. Pound and J. L. Snider. Effect of Gravity on Gamma Radiation. *Physical Review*, 140:788–803, November 1965.

[6] Peter H. Dana. The global positioning system. Geographer's Craft Project, Department of Geography, The University of Colorado at Boulder. Internet `http://www.colorado.edu/geography/gcraft/notes/gps/gps_f.html`, May 2000.

[7] Satellite Navigation and Positioning Group, University of New South Wales. About GPS. Internet `http://www.gmat.unsw.edu.au/snap/gps/about_gps.htm`, February 2002.

[8] J. C. Hafele and R. E. Keating. Around-the-World Atomic Clocks: Predicted Relativistic Time Gains. *Science*, 177:166–168, July 1972.

[9] J. C. Hafele and R. E. Keating. Around-the-World Atomic Clocks: Observed Relativistic Time Gains. *Science*, 177:168–170, July 1972.

[10] N. Ashby. Relativity in the Global Positioning System. *Living Reviews in Relativity*, 6:1–+, 2003.

[11] Trimble. *All About GPS*. Trimble Navigation Limited, Internet `http://www.trimble.com/gps/`, 2003.

[12] U.S. Dept. of Defense. Command, Control, Communications, and Intelligence, OASD(C3I)/ODASD(C3ISR&S)/Space Systems, Attn for Assistant for GPS, Positioning and Navigation, 6000 Defense Pentagon, Washington DC 20301-6000, U.S.A. *Global Positioning System Standard Positioning Service Performance Standard*, October 2000.