GPS

Professor Richard Harvey FBCS

12th October 2021

It is 01:40 on 19th July 1977 in Cedar Rapids, Iowa, three engineers at Rockwell Collins are faffing around trying to receive the weakest of weak radio signals from satellite NTS-2. At night the radio environment is often quieter, but the US Airforce has set-up a competition to be the first to decode the signal. The team are somewhat deflated to receive only a stream of "A"s. It's only later that it becomes evident that they received the first GPS signal every sent and thus a technology that transformed the world has begun.

Of course, the history of navigation via celestial objects is very long but it's not fair to draw an inventive lineage from astronomy or even Doppler. The critical feature of GPS is time-of-flight timing and it's here where the story should start. Radio waves travel at 299,792,458 meters per second in a vacuum and unlike, say, sound waves, the speed of propagation is nearly constant[1]. Thus, we could use the time it takes for a radio wave to travel somewhere, the propagation delay, as a very accurate yardstick The British Royal Airforce in World War II had a pressing need for accurate navigation – the daylight bombing raids were very expensive so taking a leaf out of the Luftwaffe's book, they decided to bomb at night. But to navigate successfully over enemy territory at night? The solution was a system known as Gee.

A typical Gee setup used three transmitters: the master (usually termed A) which sent out a blip at regular intervals that was used to trigger an oscilloscope in the aircraft; the first slave, B, which waited until it received A's blip and then sent out its blip after a 1ms wait time and a second slave, C, which listened for A's second pulse (which was sent 2ms after the first) and then sent out its blip (after waiting 1ms). The aircrew measured these time differences on an oscilloscope screen and then used a special chart to work out where they were. Gee is regarded as the first navigation system to use time-of-flight measurements (to be precise it uses time differences rather than absolute times). Later systems such as LORAN-C used the same idea[2]. Gee is interesting because of the design decisions. The transmitters were based on land, so transmitter power was not really an issue (around 300 kW was commonplace). But multiple frequencies imply a more complex receiver in the aircraft, so the system used a single carrier frequency (48.75MHz for the first chain)[3].

Gee and LORAN are sometimes known as hyperbolic navigation systems. To see why this is, consider the first Gee system with a Master transmitter at Daventry, UK and a slave at Clee Hill Shropshire. These stations are very close to 100km apart so a radio wave will take around 333 microseconds to travel between them (from A to B). If we are midway between the stations, then the pulse arrives in 167 microseconds. The second pulse arrives in 333 microseconds (AB) +167 microseconds (AB/2), so the difference is zero. Thus, if the Bomber is midway between the stations, we see zero difference in arrivals. Measuring such small differences is rather tricky with analogue electronics and requires a rather bulky device called an oscilloscope and rather remarkably WW2 bombers took off with oscilloscopes on board and the navigator was expected to read the time differences and look at a map marked with hyperbolae and deduce the location of the aircraft. Needless to say, such mental gymnastics required considerable training.

---

[1] In air the speed of light is around 56 ms-1 slower.

[2] Measuring time differences is quite expensive with analog electronics, so the British version of LORAN, Decca Navigator, used phase differences which are a lot simpler to detect.

[3] The pulses arrived at different times so this was a form of what we would now call time-division multiplexing.

In practice one needs at least one other transmission station to get an accurate fix to a Gee *chain* consisted of least three stations and four were needed in some cases.

Although Gee was complicated to use, the principle was sound and additionally it introduced the idea of a time-base generator that allowed the oscilloscope in the receiver to sync-up. The idea was adopted by the US Army Air Corps and modified slightly to allow the US Navy to use it, and by the end of WWII there were around 72 Loran stations and tens of thousands of receivers. A constant hassle, particularly for commercial and amateur navigation, where cost is a challenge, was the requirement to measure time delays. Measuring phase differences is so much easier with analog electronics but, because tones are sinusoidal it leads to ambiguous zeros and many of them. However, by using multiple frequencies, all chosen at a multiple of some base frequency, one can measure multiple phase nulls which will allow identification of one's position on a special chart. The Decca Navigator System (DNS) used frequencies of around 100kHz which equates to a wavelength of around 3000m. In good conditions and, used properly[4], both Loran-C and Decca could localize a ship to within a few hundred meters. In poor conditions that error could rise to many kilometers.

The advent of digital technology made a dramatic change – it was now easy to count things. So, counting regular pulses, which is how we measure delays, was remarkably easy. Thus, the time was ripe for a return to navigation by measuring propagation delays. But what frequency to use? Books are written about radio wave propagation but, to simplify very considerably, lower wavelengths travel a lot farther because they can diffract around the earth's curvature, but they require large antennae, much power to transmit and their long wavelengths imply low positional accuracy. High frequencies are associated with short antenna, line of sight propagation and excellent positional accuracy. The designers realized that line-of-sight propagation was of limited use to most navigational solutions – if you can see your port then you hardly need to know where you are[5]. But the advent of satellites presented an intriguing possibility -- celestial objects generally have a highly measurable and predictable positions[6] and, once overhead, satellites are clearly reachable by line-of-sight communication with high frequency radio waves.

And in 1977 the first communication was received and by 1993 a full constellation of 24 satellites was available.

So how does this exciting technology work? Well, it is simple to login into gps.gov, but it is certainly not simple to decipher the GPS standards. Part of the difficulty is that it a mixed system: part military and part civilian. Another part of the difficulty is that the system has changed over the years – some space vehicles[7] have more capabilities than others. That said, it is not really possible to understand GPS without understanding the signals it transmits. Let's focus initially on the civilian system which transmits in the L-band[8].

Each space vehicle has a very accurate clock which is in synchronization with all the other transmitter clocks. The accuracy of the clocks is further guaranteed by the ground control stations proving further synchronization with the master clock[9]. But how to measure the propagation time from satellite to the user's receiver? Gee used pulses (time-division multiplexing) but each Gee transmitter used much power and pulses are notoriously susceptible to interference from lightening, power transients and so on. The GPS designers could have used a different frequency for each satellite, but this would have meant designing broadband antennae (inefficient) or having non-interchangeable satellites. So, the designers decided that each GPS satellite would transmit on the same frequency, but each would modulate the carrier frequency using a unique code. In our simple model we are assuming that the unique code, known as the Coarse/Acquisition or C/A code in the parlance of GPS, changes the phase of the carrier wave. If the C/A code is one then we transmit one phase, if it is zero then we transmit the opposite (ie 180 degrees or an inversion). This is a modulation scheme known as Binary Phase Shift Keying and is illustrated at the bottom

[4] Needless to say, they were frequently not used properly and there is quite a bit of evidence that fishermen did not apply the appropriate DNS corrections

[5] An important exception is the business of landing an aircraft in poor visibility – a task that is considerably simplified by electronic navigation aids.

[6] Or *ephemeris* as astronomers like to call it.

[7] GPS standards refer to satellites as "space vehicles" which as the great advantage of avoiding have to spell satellite.

[8] I should note at this point that because of its US military origins the GPS system and its successors has become a morass of impenetrable acronyms. Some GPS primers feel compelled to include a multi-page glossary. I will just do my best to keep the alphabet to a soupçon rather than a soup.

[9] If we were just computing location, then there would be no need for the GPS system to keep a record of absolute time but users like it so the GPS system also transmits a conversion to UTC.

of Figure 1. The carrier is an exact multiple of the GPS atomic clock, and all the other codes are likewise at exact multiples so, once the receiver has synchronized, the receiver has a version of a very accurate clock.

The C/A code is used to firstly identify the satellite, but also to determine the propagation delay and hence the distance to the satellite. The C/A code consists of 1023 bits (usually called *chips*) and repeats every 1 ms. Each code is different but looks random – a pseudo-random sequence. Furthermore, by selecting the codes from a family known as Gold codes, we can be sure that each code does not match another. The receiver keeps a copy of all the possible 37 C/A codes, and it tries to match each possible code with the incoming signal. Because there is an unknown propagation delay it tries every code with every possible delay (delays are usually quantized in half chips). Once the receiver gets a match, we have a synchronization in the receiver then two important things have been discovered – the GPS clock and the propagation delay. This has to be repeated for at least four satellites for a "fix".

An important feature of Gold codes is that once we are synced up to a satellite, the Gold codes from the other satellites just look like noise – a it's a system known as Code Division Multiple Access or CDMA. It allows for very low power transmission which is helpful because power is very restricted on a satellite (the minimum signal power is around 22W) but as we add more satellites the signal-to-noise ratio at the receiver drops (each satellite transmits its own code which is out of sync with the other codes so looks like noise). As a consequence, satellite transmission powers have tended to increase and/or code lengths have increased.

The original military GPS introduced a second frequency, known as L2 which transmitted a much longer code known as the P code[10].

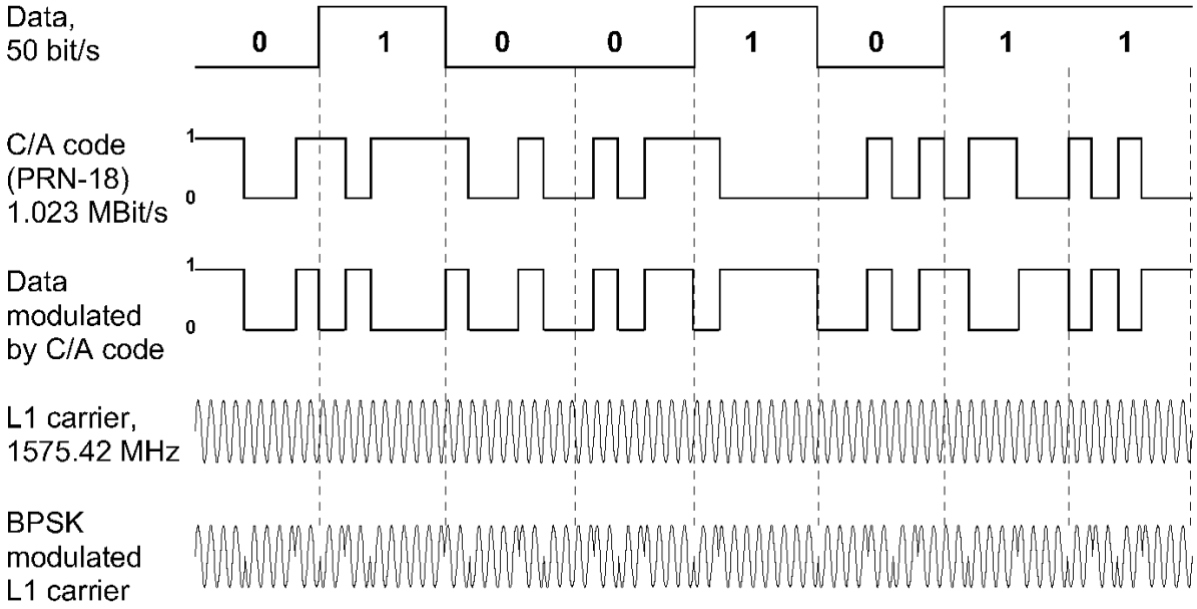

Figure 1: *Showing simplified GPS signals – taken from Figure 42 of [1]*

The final bit of the puzzle is that the satellite also needs to transmit some side-data to either improve the accuracy of the positional estimate or for user convenience. Without these side-data the GPS receiver can only estimate a *pseudo-range* which is the uncorrected delay. These are known collectively as the Navigation message which includes, among other things, time synchronization signals, ephemeris for the satellite, orbital data for the satellite (almanac), correction signals for propagation time (ionosphere corrections for example) and some information on the operating health of the satellite.

All of which should make it clear that a GPS receiver is a very complicated thing. At the 'front-end" it contains

[10] Originally this was transmitted without encryption, but in 1994 this was encrypted and is referred to as the P(Y) code. It's the same idea but with a longer code running at a faster rate to give better positional accuracy.

an antenna designed to receive the carrier wave[11] a radio receiver and an accurate reference oscillator which has to be synced-up to the GPS clock. Following that is a signal processing section which laboriously tries every Gold code at every possible delay for every satellite to estimate the propagation delays and then at the end of the chain is computer which applies all the corrections and computes the location of the antenna. It is a complete miracle of miniaturization that all of that can be compressed into a module about the size of a pencil tip.

In terms of the history of the technology the first GPS receiver was about the size of three euro-cabinets but was later miniaturized to something that the army call a "manpack". This is a picture of one in the Smithsonian Museum. It cost about $45,000 and weighs around 9kg. Needless to say, it was not that popular with the soldiers that were asked to carry it. In terms of the satellites, these are quite heavy, around 1.5 to 2.0 tonnes since, as they operate near the van Allen belts there is need to shield the interior electronics. They have a design life of 10 years and in the early years used around 800W of power but, more recently, that has been increased to a couple of kW.

Now I'd like to welcome you to the fiendish world of US military acronyms. GPS technologies are divided into "Blocks" which are generations of satellites. Thus, there are currently eight satellites still operation from Block IIR. For manufacturers and system designers there is problem known as "backwards compatibility". How can we ensure that receivers made in the 1990s still function? The early decision to use CDMA or spread spectrum[12] is a big help since more signals can be crammed onto a single carrier by adding more codes (which look like increased noise which in turn can be countered by more directional antennae or more power) but it makes for a complicated signal structure. Compared to the original two signals transmitted on two carriers a modern GPS satellite transmits six signals on three bands using a fiendishly clever phase modulation scheme that is backward compatible with BPSK but has better spectral characteristics[13].

Undoubtedly GPS was the first and most widely celebrated of the satellite based hyperbolic navigation systems but there are others. Together they form the Global Navigation Satellite System or GNSS. The most venerable is GLONASS which is a Russian system. Some of the GLONASS satellites have high orbital angles (around 65 degrees from the equator) which makes for better reception in polar regions. Unlike GPS, GLONASS has each satellite transmitting the same code but on a different frequency – it uses a Frequency Division Multiple Access (FDMA) system[14]. Galileo and Beidou have stuck to CDMA. Galileo is a worldwide system, but Beidou covers only Asia. Given its regional coverage, Beidou, was also the first system to use geostationary satellites, although the Indian system, Navic now also has three geostationary satellites above India. It is worth mentioning at this point that the expensive decision to construct a GNSS network often arrives from a complex balancing of political interests and alliances[15]. Some countries feel that depending on another country's navigation system is too big a risk to take. A compromise position is that adopted by Japan with the Michibiki system.

The Michibiki system, also known as QZSS is an augmentation to GPS. The satellites have unusual elliptical orbits so overcome the problem of lack of GPS satellite visibility in Japanese urban areas (tall buildings) and deep valleys in rural ones. But the system also uses ground stations to measure errors in the GPS co-ordinates and transmit corrections either from the Michibiki satellites or from ground stations.

The idea of augmenting GPS is not a new one and arose from an earlier version of GPS when the military had the potential to deliberately degrade the civilian signal via a process known as "selective availability".

---

[11] The usual rule of thumb is that an effective antenna will have dimensions of around a wavelength if it is to be directional and offer any gain. The wavelength of L1 is around 0.3m so obviously most receivers have omnidirectional antennae with no gain.

[12] I've avoided using the term spread spectrum up till now, but I see it has slipped in. Spread spectrum also includes a modulation technique known as frequency hopping in which the transmitter changes the carrier frequency at regular intervals. I'd like to be able to add a footnote to a footnote but let me just briefly state that neither frequency hopping not spread spectrum were invented by Hedy Lamarr.

[13] Enthusiastic readers are referred to Binary Offset Code Modulation (BOC) and multiplexed BOC (MBOC).

[14] The updated GLONASS is reputed to be moving to a CDMA system though.

[15] It is stated frequently on the web, and in newspapers, that during the 1999 Kargil war between Pakistan and India that the US denied India access to GPS data. This seems improbable to me. As we have seen, the civilian use has been available since the 1980s. Possibly the Indian government asked the US to turn-off a feature called "selective availability" which was a corruption that the US military used to apply to the civilian signal but there is no evidence that anything was asked for, nor denied.

One of the solutions was set up ground stations which received the GPS signals and transmitted, usually via terrestrial, radio corrections known as differentials. In the UK differential GPS never really caught on for maritime use and the three General Lighthouse Authorities of UK and Ireland will discontinue the service next year. Furthermore, the current GPS satellites have no selective availability, and better receivers and tracking have led to excellent positional accuracies.

For aircraft, the situation is very different, and a number of authorities have proposed satellite-based augmentation systems: WAAS in the USA, EGNOS in Europe. The idea is that ground stations measure errors which are then beamed up to geostationary satellites which then broadcast corrections that suitably equipped devices can use. Alternatively, there may be an internet data service (EDAS in Europe) that allows portable devices to apply appropriate corrections[16].

Leaving Brexit to one side, one might ask the extent to which the GNSS can be relied upon? There are essentially four failure modes associated with GNSS. The first two result from deliberate attempts to corrupt the signal: jamming and spoofing. The third results from adverse weather and the fourth, space warfare, is a new field and is at the point where most super-powers acknowledge it could be a domain of warfare[17] without being very clear what form combat will take.

Jamming and spoofing, however, are current threats. Jamming is the blocking of the signal so that receivers cannot determine their position. The ease of jamming arises from the low signal power at the receiver and as previously mentioned, many GPS receivers are omni-directional. Thus, a powerful ground-based transmitter[18] at, say, L1 can easily overwhelm a receiver's front-end. Recent work on the International Space Station [2] has found an example on a Syrian airbase – an 80W ground-based transmitter that appears to be transmitting C/A codes 1 to 32 at L1 (a particularly potent form of jamming since it confuses the signal-processing section of the receiver) and at L2 just a narrowband signal.

Spoofing is even trickier to detect but there have been several reports from ships captains of their GPS systems reporting false positions around Russia and the Crimea. Enterprising Defence Analysts at the Center for Advanced Defense Studies (C4ADS) in Washington DC have correlated these disturbances with the appearance of President Putin in sensitive areas. While the link may not be causal, it is true that software defined radios (SDRs), which are the essential component of such systems [3], are now cheap to obtain. It is also very easy to spoof one's own position as can be attested by the number of cheats on virtual reality games such as Pokémon Go.

Just as the benefits of the GNSS have been rightly lauded with the system winning a Queens Award for Engineering in 2019, the costs of failure are now grimly apparent (see [4] for an example calculation for the UK and [5] for a technical impact statement). There would clearly be very significant impacts on the transport sectors but also on the emergency services and justice systems which use the accurate timing and location information for critical activities[19].

Space weather is also known to cause unavailability of certain components of the GNSS – WAAS was unavailable for 30 hours in October 2003 due to Ionospheric disruptions and large-scale infrequent disruptions such as the Carrington event of 1859 may well be serious.

In the UK, these concerns were acknowledged by the Cabinet Office in 2017 with the then Minister, Caroline Noakes, writing that it was important that an alternative to GNSS, such as eLoran, should be established. That letter was written two years after eLoran was discontinued by General Lighthouse Authorities. Since then, the UK has withdrawn from the Galileo and EGNOS programmes, so it is now in the unique position of being the only permanent member of the UN security council without any controlling rights over the GNSS. However, other countries have taken great care to secure their access and control of the GNSS as it is now seen as, like the internet, one of the indispensable services of the modern age.

[16] EDAS is no longer available in the UK post Brexit. And the Civil Aviation Authority has instructed pilots that they should no longer use enhanced GPS approaches via EGNOS. For large airports there are alternative systems. For small airports there are no alternatives and the banning of EGNOS is a degradation in safety.

[17] NATO defined space as the fifth operational domain, alongside air, land, sea and cyberspace, in December 2019.

[18] Actually, the transmitter need not be that be powerful. Trinity House reported experiments on one of its vessels that corrupted multiple GPS onboard systems which reported false positions without apparently detecting anything wrong. The jammer was less powerful than a mobile phone.

[19] [5] also presents informal experiments which imply that GNSS jamming and/or interference is commonplace.

Bibliography

[1] J.-M. Zogg, "GPS- Essentials of Satellite Navigation Compendium, report GPS-X-02007-D," U-Blox AG, 2009.

[2] M. J. Murrian, L. Narula, P. A. Iannucci, S. Budzien, B. W. O'Hanlon, M. L. Psiaki and T. E. Humphreys, "First results from three years of GNSS Interference Monitoring from Low Earth Orbit," *NAVIGATION: Journal of the Institute of Navigation,* p. Under Review, 2021.

[3] W. Feng, J.-M. Friedt, G. Goavec-Merou and F. Meyer, "Software-defined radio implemented GPS spoofing and its computationally efficient detection and supression," *IEEE Aerospace and Electronic Systems Magazine,* pp. 36--53, 1 March 2021.

[4] G. Sadler, R. Flytkjaer, F. Sabri and D. Herr, "The economic impact on the UK of a disruption to GNSS," London Economics for Innovate UK, London, 2017.

[5] C. Witty and M. Wolpert, "Satellite-derived Time and Position" A Study of Critical Dependencies: Blackett review," Government Office for Science, London, 2018.

[6] C. Dickens, Great Expectations, London: Chapman and Hall, 1859.

[7] J. Smith, ""How to write an effective transcript"," *Writing Today,* pp. 34-56., 2011.