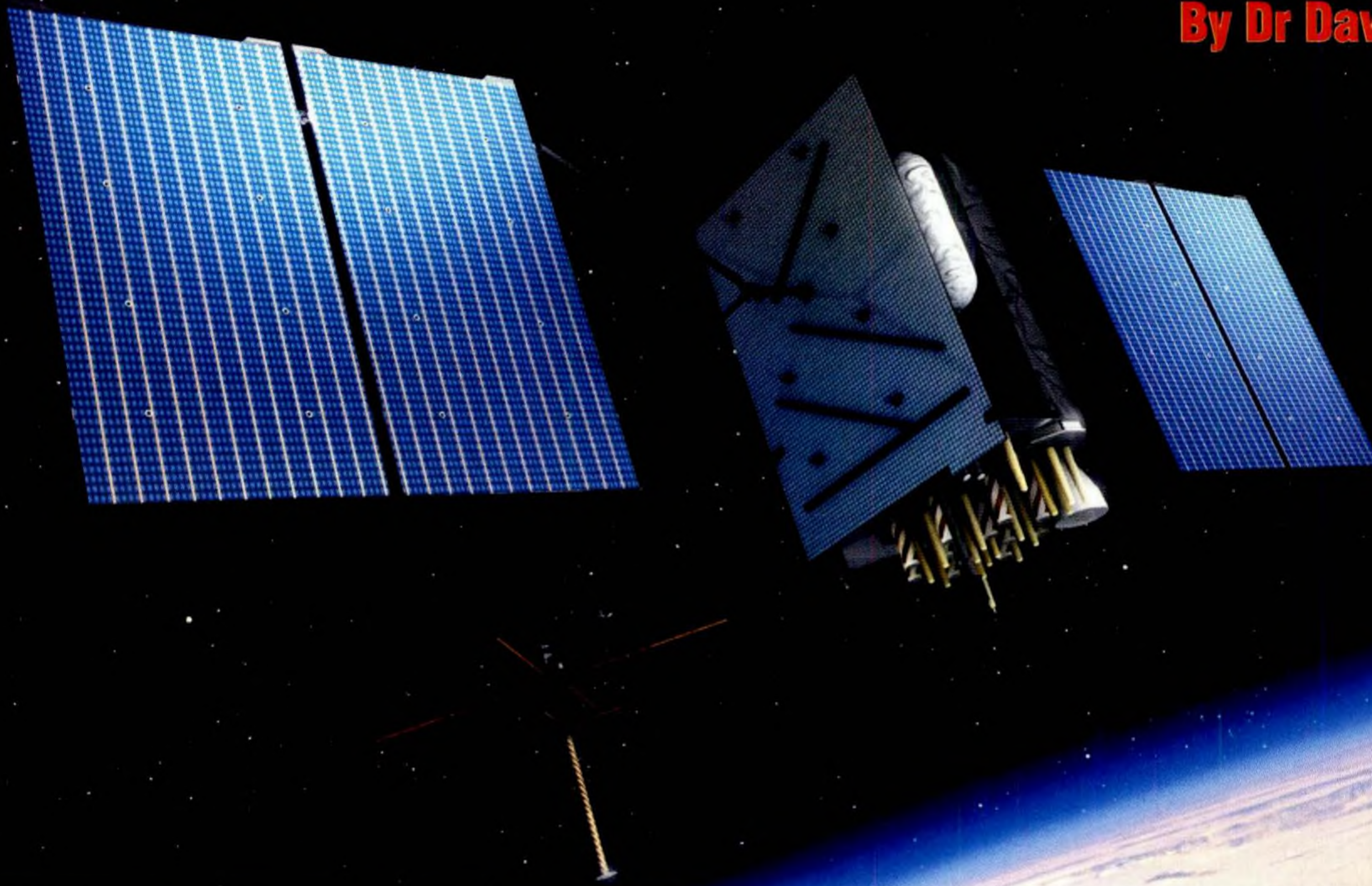


How does satellite navigation work?

A look at GNSS systems

Global Navigation Satellite Systems (GNSS), of which the familiar GPS (Global Positioning System) is but one example, are now ubiquitous and used in everyone's daily lives. Most people do not even know the origin or workings of the location and mapping functions built into the smartphones that they carry. To them, it's just 'there', and it works. But there is a lot going on behind the scenes!

By Dr David Maddison



In the September 2018 issue of SILICON CHIP, we published an article on Augmented GNSS (siliconchip.com.au/Article/11222), describing how the accuracy of satellite navigation systems can be enhanced beyond what is ordinarily available, through various augmentation systems (eg, SBAS – Satellite-Based Augmentation System).

This augmentation is not needed for ordinary users but is for applications such as aircraft landing, precision agriculture and self-driving cars etc.

We also looked at a predecessor system to GPS, the terrestrial based Omega Navigation System in the September 2014 issue (siliconchip.com.au/Article/8002).

But so far, we have not actually described in detail how satellite navigation systems work. This article corrects that omission. We will go back to basics, to describe how the regular (non-augmented) GNSS systems operate.

In the beginning . . . the word was the US GPS

The first GNSS system put in place, and the one most people are familiar with, is the US Global Positioning Sys-

tem (GPS). This was developed for the US military, both for navigation and to ensure better accuracy with their weapons systems (with the benefit of limiting unintended civilian casualties).

This system was also made available free of charge worldwide, with reduced accuracy at first ("selective availability"), then later with the full available accuracy. The USA turned off selective availability on 1st May 2000 and undertook never to use it in the future.

Part of the motivation for this was to prevent future tragedies such as Korean Air Lines Flight 007, which was shot down by the Soviet Union in 1983 after inadvertently flying into Soviet air space, due to a navigational error.

Newer GPS satellites, or Space Vehicles (SVs) as they are called, don't even support selective availability.

While the US Global Positioning System was the first, the following systems have since come into service, or soon will be: GLONASS (Russian; fully operational), Galileo (EU, to be fully operational by 2020) and BeiDou (China, also to be fully operational by 2020). There are also two regional

systems: NavIC (India) and QZSS (Japan).

Then there are the following augmentation systems which provide greater positional accuracy (down to cm or even mm) and which were described in the September 2018 article: WAAS (USA), EGNOS (EU), MSAS (Japan), GAGAN (India), SDCM (Russia), WAGE (US Military), SBAS (Australia, test-bed) and the commercial systems StarFire, C-Nav, Startfix, and OmniSTAR.

Newer satellite navigation receivers are 'multi-constellation' types which support more than one of the above GNSS systems and can have access to over 100 satellites.

This makes position fixes in "urban canyons" and challenging terrain easier, as there is a higher likelihood of having more satellites visible directly overhead, also avoiding multi-path reflections from satellites that are not directly in the line of sight.

Basic operating principles

The same basic operational principles apply to all GNSS systems. Each system has a group of satellites in orbit, known as a constellation. Each satellite sends a continuous signal to Earth which contains data such as the satellite ID, the current time onboard the satellite, the position of the satellite and other data. For GPS, this encoded information is called the Navigation Message.

All of the satellites in a constellation are synchronised with the same time reference, which is achieved using extremely accurate atomic clocks onboard each satellite and on the ground.

To achieve a full position fix, in theory three satellite signals at sea level are sufficient (where sea level represents the roughly spherical shape of the Earth, the so-called reference ellipsoid or 'geoid' which are accurate models of the exact shape). Four satellites are required to also compute altitude above sea level.

To get a position fix, two fundamental things need to be established. The first is the distance from the user's receiver to three, four or preferably more satellites.

This is called "trilateration" in the specific case of three satellites, or "multilateration" for three or more.

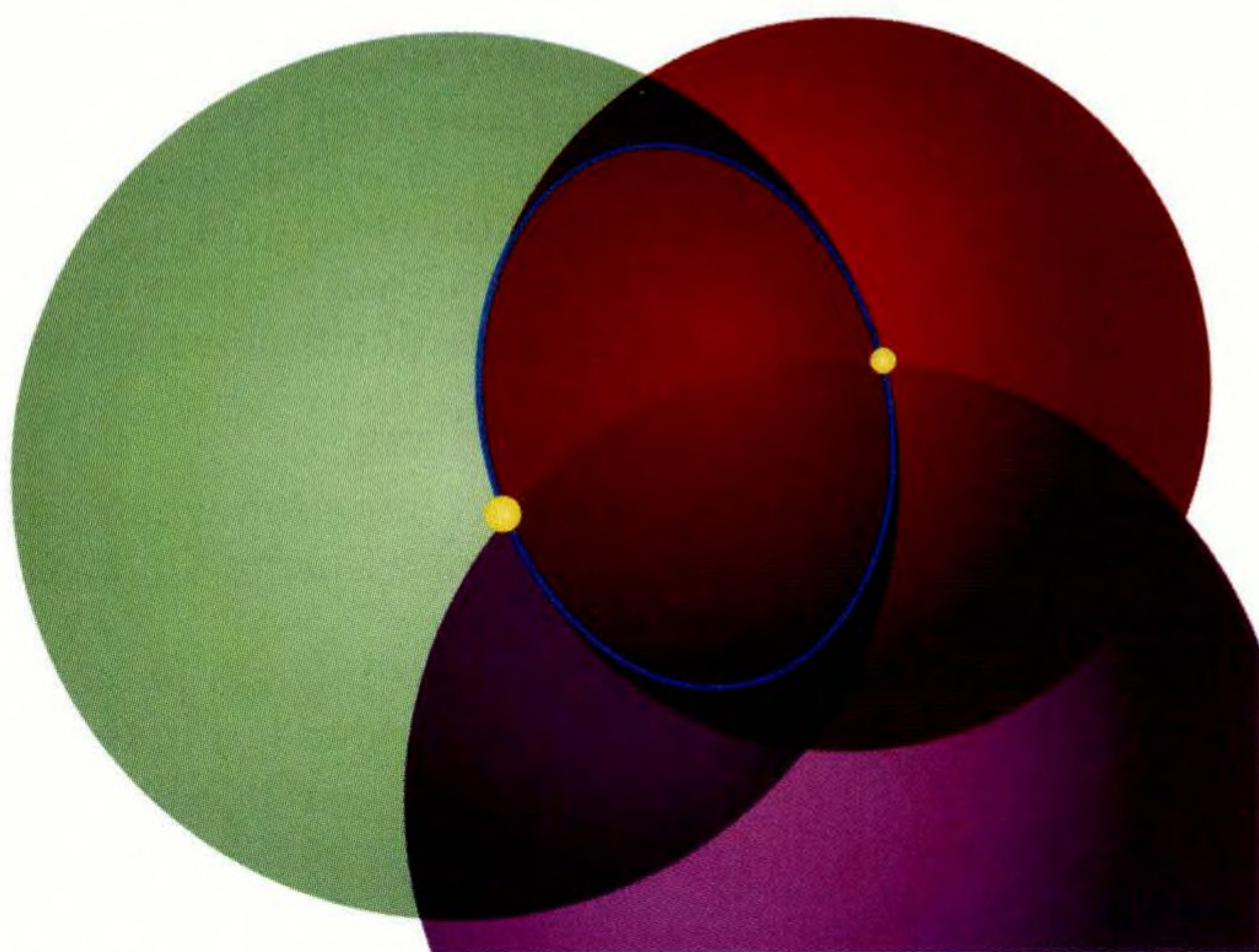


Fig.1: the intersection of three spheres, with radii defined by the distance between a group of satellites and a receiver. This shows how the intersection of two spheres produces a circle (blue), and the addition of a third sphere defines two points on that circle (yellow).



(Above and opposite): an artist's impression of the latest generation GPS Block IIIA satellite by Lockheed Martin, first launched December 23, 2018. These offer three times greater positioning accuracy than their predecessors, increased signal power and much-improved resistance against jamming. The satellites of the GPS constellation are named NAVSTAR (Navigation Satellite Time and Ranging) with various numbers to identify them. See the video titled "Building the Most Powerful GPS Satellite Ever - GPS III" at: siliconchip.com.au/link/aavj

This gives a relative position of the receiver with respect to those satellites at the time of their transmission. To calculate the user's location, it's therefore also necessary to establish the position of the satellites at the time they transmitted their signals, which is encoded in the data stream along with the time of transmission.

This then gives the approximate location of the receiver on the Earth's surface or above it. These measurements are then followed by many corrections and iterative adjustments to get a more exact positional fix.

Determining the distance to the satellites

Radio signals travel at the speed of light, ie, 299,792,458m/s

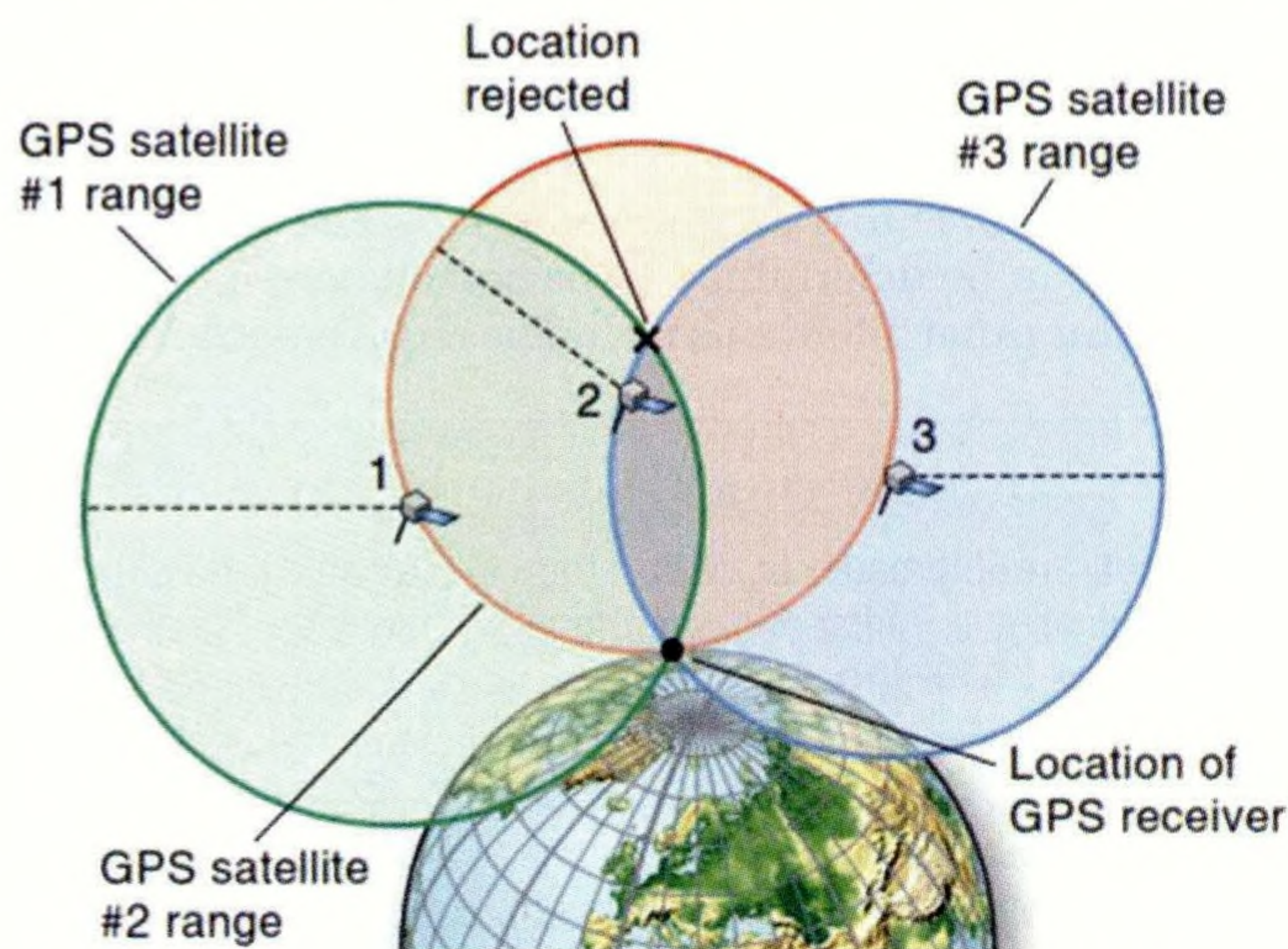


Fig.2: the intersection of spheres representing the distance from a receiver to three satellites, showing the two possible locations of the receiver with one point being obviously wrong and rejected. A fourth satellite will establish additional information such as altitude and help in calculations to correct the receiver time clock.

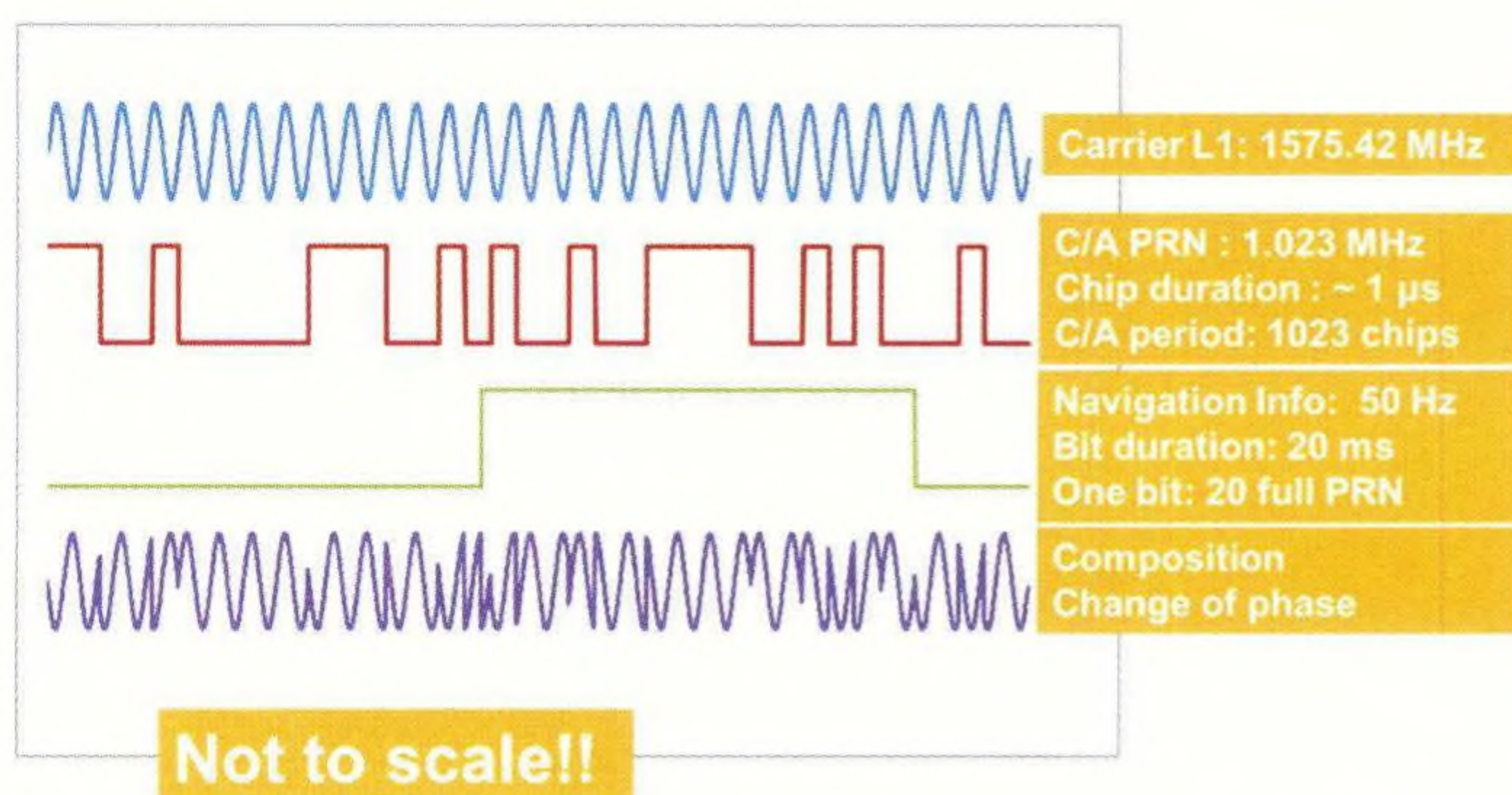


Fig.3: the structure of GPS signals including a carrier, pseudo-random noise (PRN) code, navigation data (one bit for every 20 PRN cycles) and the resulting combined signal, that is transmitted by the satellite. C/A stands for coarse/acquisition code. Image source: José Caro Ramón, Head of GNSS Augmentation Systems and Services at GMV, PMP; Creative Commons Attribution-Share Alike 3.0 Unported license.

in a vacuum. If we know the delay between the transmission of a signal from a satellite and it being received on Earth, we can determine the distance between the transmitter and the receiver.

This propagation delay calculation requires that the time the signal leaves the satellite and the receipt time at the receiver be known. The signals leaving the satellite have a time stamp of the departure time.

Ideally, the receiver would have an atomic clock synchronised to the same time as the satellite clock, but generally, this is not the case; not everybody has pockets large enough to carry around an atomic clock, or the batteries required to run it!

We will discuss how that problem is dealt with later.

Knowing the exact time is essential as even a 1ns (one-billionth of a second) clock error at the receiver compared to the satellite will result in a 30cm positional error; that is how far radio waves travel in 1ns.

Knowledge of the propagation time of a signal (ie, distance) from one satellite to a receiver locates the receiver on a sphere around the satellite, with its radius being the calculated distance (see Figs.1 & 2). Knowing you are somewhere on a sphere is not that useful, so more information is required.

If the distance to a second satellite is known, then the receiver can be determined to be somewhere in a sphere surrounding that satellite as well. The receiver location is on

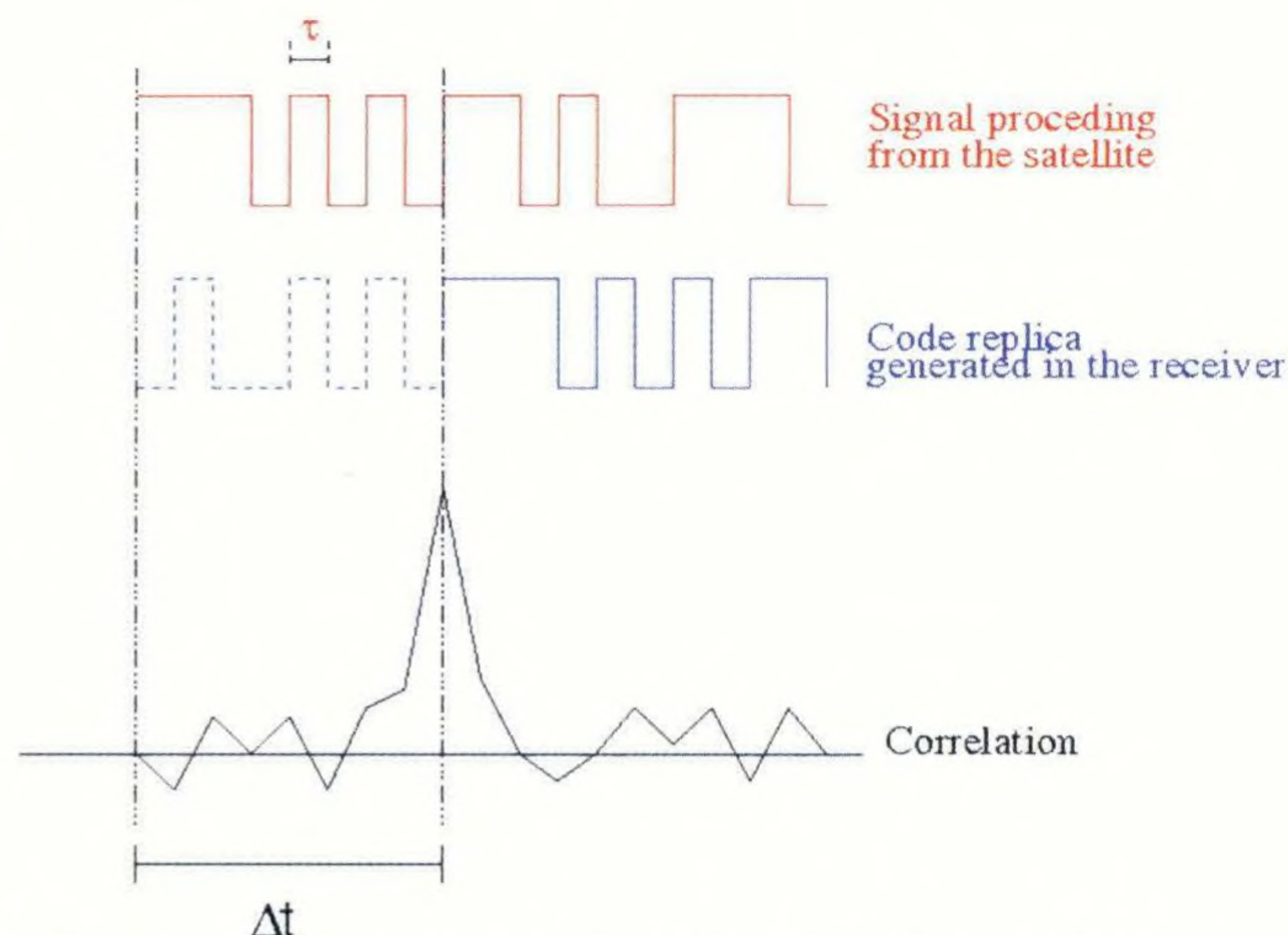


Fig.4: how the signal travel time from the satellite to the receiver is determined. The receiver knows the matching PRN code expected from each satellite. The PRN code within the receiver is shifted until the maximum correlation is found between the received and expected signals, and the offset required to do this establishes the travel time, Δt . In this case, if the receiver code was shifted all the way to the left, there would be an exact correlation. Figure source: Carlos Lopez, https://gssc.esa.int/navipedia/index.php/GNSS_Basic_Observables

the intersection of two spheres, which describes a circle. So we know the receiver is somewhere on that circle. But that is still not enough information for a complete position fix.

With a third satellite, we calculate a third sphere, and this also has to intersect with the circle formed by the intersection of the first two spheres. When a sphere intersects with a circle, it does so at two points.

So with three satellites we then have two possible positions of the receiver. The question then is how to determine which of those positions is the actual location.

Usually, the position nearest Earth would be chosen as the obvious location, and the second position would be rejected.

If a fourth satellite is used, it can unambiguously establish which of the two possible positions is the correct one without having to guess. The fourth satellite is needed for another reason as well as will be discussed later.

Note that at this point, only the relative position of the receiver with respect to the satellites is known. So to determine the actual position of the receiver with respect to the Earth, knowledge of the satellites' position is required.

Relativity effects and corrections

Satellite navigation is an everyday situation where Einstein's theories of Special Relativity and General Relativity have to be taken into account.

Firstly, because the satellites are moving relative to the observer (about 14,000km/h for GPS), there is a time dilation effect. Special relativity says that the clock on board the satellite will fall behind ground-based clocks by about 7μs per day. Bearing in mind radio waves travel about 30cm per nanosecond, this would amount to an error of 2.1km per day.

Secondly, massive bodies such as the Earth distort space-time and the closer to such a body a clock is, the slower time seems to go relative to an outside observer.

Since the satellites are high above the Earth, an observer

The full GPS interface specification

If you are interested in seeing the core technical document that defines everything you need to know about the "interface" between the "space segment" of the Global Positioning System and the "user segment", some of it is contained in the 224-page document named "Interface Specification IS-GPS-200J, May 22, 2018", available at: siliconchip.com.au/link/aavk

This describes the structure and content of data transmitted from GPS satellites on radio frequency links L1 and L2. Related technical documents can be found at: www.gps.gov/technical/

A useful book on GPS is P. Misra and P. Enge, Global Positioning System: Signals, Measurements and Performance, Ganga-Jamuna Press, 2011.

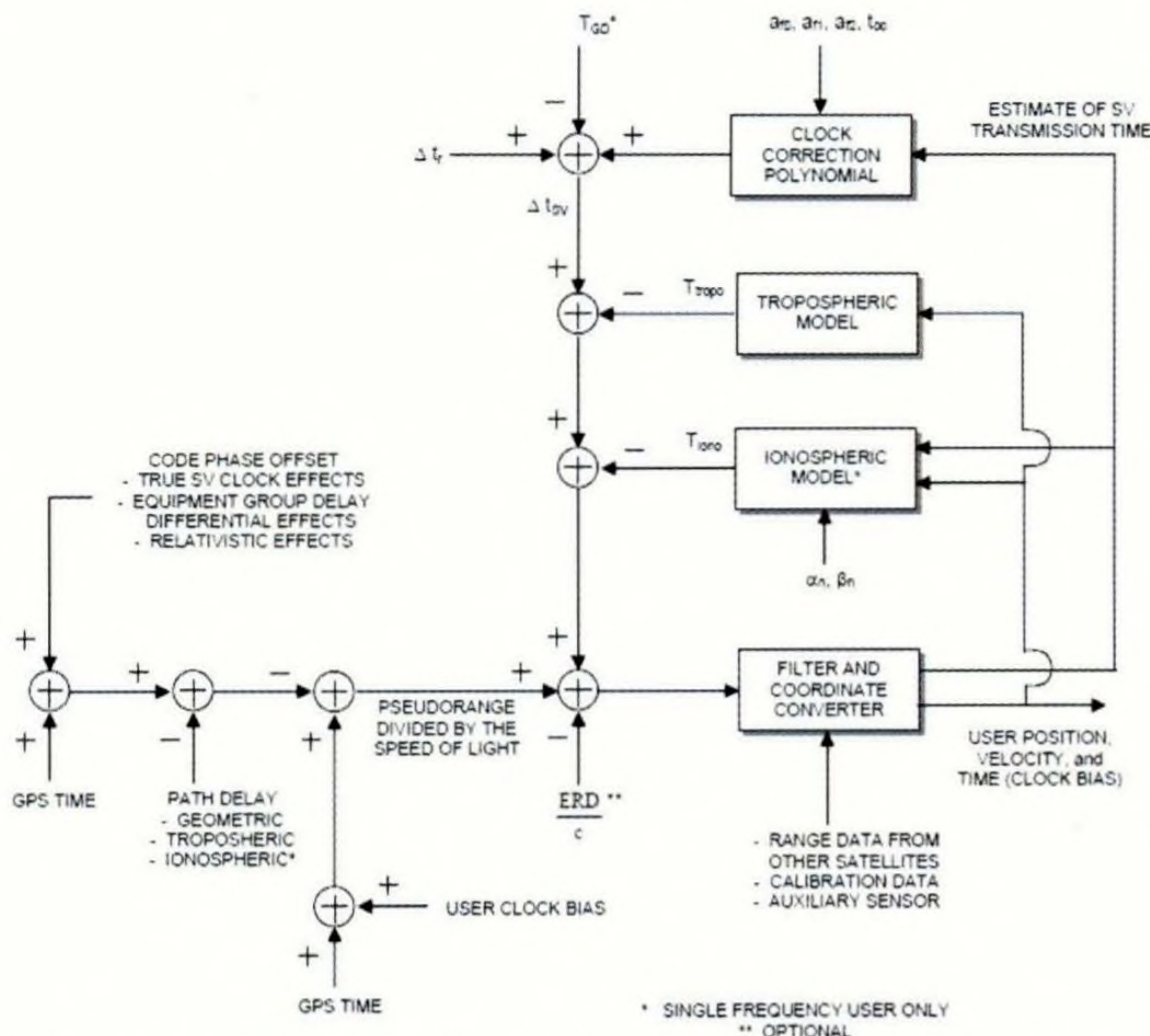


Fig.5: the mathematical model used by a GPS receiver to apply correction parameters. Similar procedures apply to other GNSS systems. SV stands for space vehicle; a_{10} , a_{11} and a_{12} are polynomial coefficients related to satellite clock error; t_{oc} is "time of clock"; ERD is estimated range deviation; c is the speed of light; t is the true GPS time at the time of data transmission; t_{sv} is the space vehicle time; Δt_{sv} is the difference between the space vehicle time and the centre of its antennae; Δt_r is the relativistic correction; α and β are ionospheric parameters; T_{GD} is the group delay differential; and T_{tropo} and T_{iono} are corrections for tropospheric and ionospheric delays. From Interface Specification IS-GPS-200K, "NAVSTAR GPS Space Segment/Navigation User Segment Interfaces".

on Earth would see the satellite clock running faster than an Earth-based clock by about 45 μ s per day.

The combined effect of the satellite clock running slower due to special relativity and faster due to general relativity from the point of view of an Earth-based observer is a difference of 38 μ s or 11.4km per day. Satellite navigation would therefore be worthlessly inaccurate if these relativistic effects were not taken into account

Another phenomenon that has to be taken into account is the kinematic "Sagnac effect". This can amount to a timing error of up to 207ns or up to 62m per day.

Between the satellite and the Earth, there is a rotating frame of reference. Two electromagnetic beams going in opposite directions on the same closed path around a rotating object will take different times to complete the trip. Therefore, the timing has to be adjusted to obtain the exact propagation time of a signal from the satellite to the receiver.

There are additional corrections which must be made to get accurate results, which will be discussed later.

The pseudo-random noise (PRN) ranging code

The pseudo-random noise code is what is used to identify which signals come from which satellites. All satellites in a GNSS constellation are assigned a unique PRN number.

In the case of GPS, two primary frequencies are used (with more under development). These are L1 and L2. Civilian GPS mostly uses just L1 (and some L2) and the military use both L1 and L2.

```
***** Week 512 almanac for PRN-24 *****
ID: 24
Health: 000
Eccentricity: 0.6661891937E-002
Time of Applicability(s): 319488.0000
Orbital Inclination(rad): 0.9505615234
Rate of Right Ascen(r/s): -0.7850758266E-008
SQRT(A) (m 1/2): 5153.627930
Right Ascen at Week(rad): -0.7220151424E-000
Argument of Perigee(rad): -0.651554227
Mean Anom(rad): -0.5415329933E-000
Af0(s): 0.1974105835E-003
Af1(s/s): 0.3637978807E-011 week: 512
```

Fig.6: example almanac data for one satellite. Each GPS satellite transmits almanac data for all GPS satellites. This allows a receiver to determine which satellites are likely to be in view, significantly reducing the amount of signal searching that it needs to do.

```
EPHEMERIS FOR SATELLITE 24 :
PRN number for data ..... 24
Issue of ephemeris data ..... 179
Semi-Major Axis (meters) ..... 2.65599E+07
C(ic) (rad) ..... -1.02445E-07
C(is) (rad) ..... -1.22935E-07
C(rc) (meters) ..... 168.656
C(rs) (meters) ..... -63.3125
C(uc) (rad) ..... -3.48687E-06
C(us) (rad) ..... 1.1526E-05
Mean motion difference (rad/sec) ..... 3.94802E-09
Eccentricity (dimensionless) ..... 0.00623617
Rate of inclination angle (rad/sec) .. 1.05004E-10
Inclination angle @ ref. time (rad) .. 0.976756
Mean Anomaly at reference time (rad) . 1.79689
Corrected Mean Motion (rad/sec) ..... 0.000145861
Computed Mean Motion (rad/sec) ..... 0.000145858
Argument of perigee (rad) ..... -2.06498
Rate of right ascension (rad/sec) .... -7.67032E-09
Right ascension @ ref time (rad) ..... -2.4059
Sqrt (1 - e^2) ..... 0.999981
Sqr root semi-major axis, (m^1/2) .... 5153.63
Reference time ephemeris (sec) ..... 252000
```

Fig.7: an example of GPS satellite ephemeris data, broadcast from each satellite. PRN is the pseudo-random noise number. The ephemeris is highly accurate orbital data from which the exact location of the satellite can be established.

In the civilian case, since all satellites are broadcasting on the same frequency, a way is needed to identify the signal from each individual satellite from among a whole jumble of signals.

The GPS date rollover problem

GPS time uses week numbers which started counting at midnight on 5th January 1980 and are numbered from 0 to 1023 (ie, 1024 weeks), after which the week number is reset to zero. The first rollover occurred on 21st August 1999, and the next one after that was on midnight 6th April 2019.

The next rollover will occur at midnight on 2nd November 2038. This year, there was a concern that some GPS units might not handle the rollover correctly and would reset themselves to 1980 or 1999. People were warned about this, but it appears to have not been a problem as most GPS units were programmed correctly to handle it.

Editor's note: we noticed some older GPS modules giving incorrect dates after April 6. Apart from the date being wrong (nearly 20 years earlier than it should be), everything else seems to work, including location information and the time. These modules were purchased some years ago; those sold within the last few years should handle the week rollover seamlessly.

COMPONENTS OF EPHEMERIS DATA

Name	Description	Units
M_0	Mean anomaly at reference time	Semicircle
Δn	Mean motion difference from computed value	Semicircle/s
e	Eccentricity	Dimensionless
$\sqrt{a}$	Square root of semimajor axis	$m^{1/2}$
Ω_0	Longitude of ascending node of orbit plane at weekly epoch	Semicircle
i_0	Inclination angle at reference time	Semicircle
ω	Argument of perigee	Semicircle
$\dot{\Omega}$	Rate of right ascension	Semicircle/s
$IDOT$	Rate of inclination angle	Semicircle/s
C_{uc}	Amplitude of cosine harmonic correction term to the argument of latitude	Rad
C_{us}	Amplitude of sine harmonic correction term to the argument of latitude	Rad
C_{rc}	Amplitude of cosine harmonic correction term to the orbit radius	m
C_{rs}	Amplitude of sine harmonic correction term to the orbit radius	m
C_{ic}	Amplitude of cosine harmonic correction term to the angle of inclination	Rad
C_{is}	Amplitude of sine harmonic correction term to the angle of inclination	Rad
t_{oe}	Ephemeris reference time	s
$IODE$	Issue of data, ephemeris	Dimensionless

Parameter	Explanation
t_{oe}	Ephemerides reference epoch in seconds within the week
$\sqrt{a}$	Square root of semi-major axis
e	Eccentricity
M_0	Mean anomaly at reference epoch
ω	Argument of perigee
i_0	Inclination at reference epoch
Ω_0	Longitude of ascending node at the beginning of the week
Δn	Mean motion difference
$\dot{i}$	Rate of inclination angle
$\dot{\Omega}$	Rate of node's right ascension
C_{uc}, C_{us}	Latitude argument correction
C_{rc}, C_{rs}	Orbital radius correction
C_{ic}, C_{is}	Inclination correction
a_0	SV clock offset
a_1	SV clock drift
a_2	SV clock drift rate

Fig.8(a) [left]: the values within the ephemeris (orbital) data and their meanings. In addition to the ephemeris, the Navigation Message also contains the following important clock parameters: t_{oc} (reference time) and a_0, a_1, a_2 (polynomial coefficients for clock correction: bias [s], drift [s/s], and drift rate/aging [s/s²]). Fig.8(b) [above] explains the symbols of Figs.8(a) and Fig.9

Fig.9 (opposite): for those interested in the mathematics behind calculating the satellite position using ephemeris data, here are the equations used. WGS84 is the World Geodetic System 1984 coordinate system, and ECEF is Earth-centred, Earth-fixed coordinate system. Table from Ryan Monaghan. From: siliconchip.com.au/link/aav1

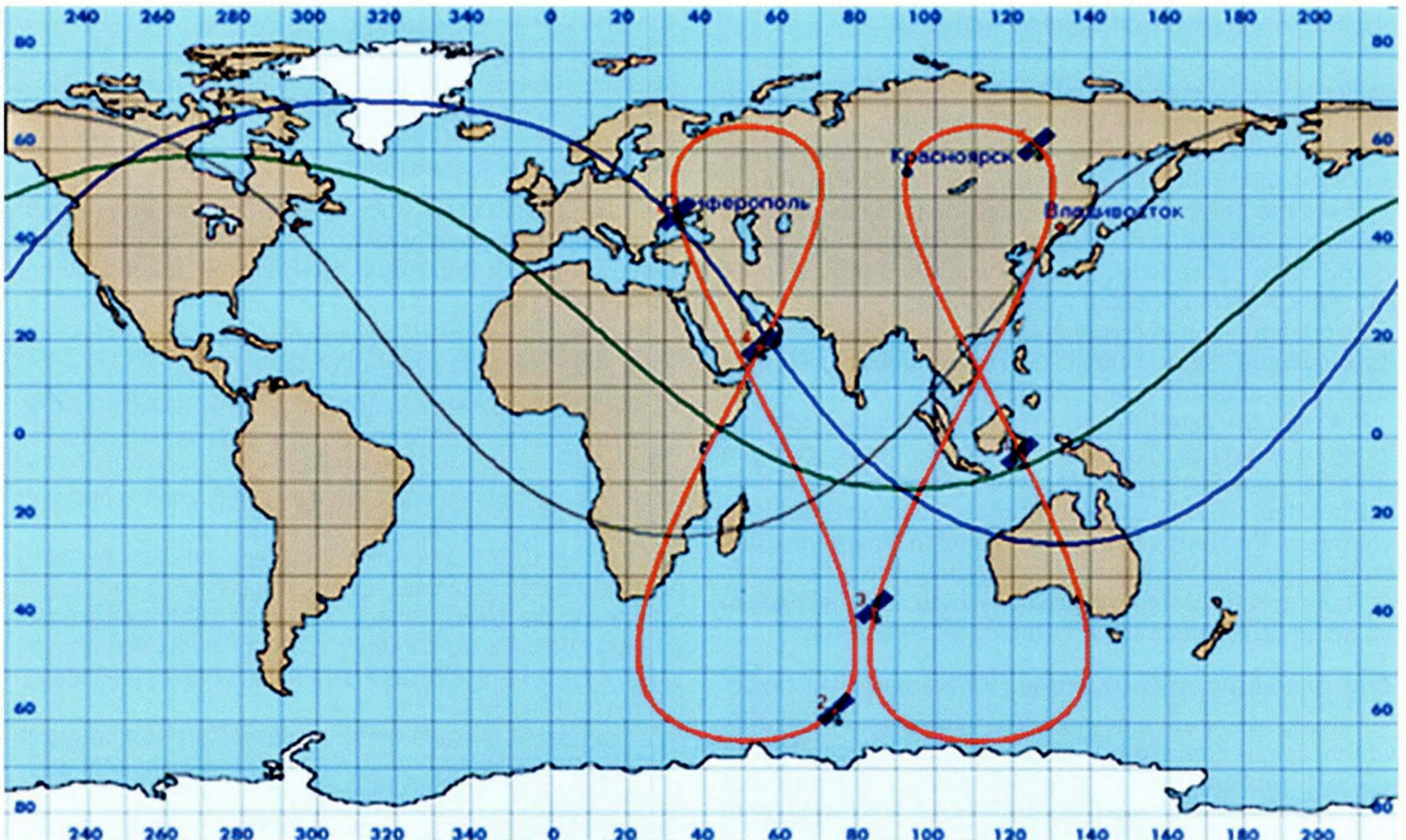


Fig.11: high-orbit GLONASS is a system that will be implemented to provide improved regional coverage over Russia, much like the Japanese QZSS system. The ground tracks of the orbits are shown in red. Presumably, Australian users will benefit from this system as with QZSS, as some of the satellites will be visible over Australia.

the voices of all other speakers are rejected as noise (different PRN codes).

In the case of GPS, the PRN code is 1023 bits long and repeats every millisecond for civilian users. The military L1 and L2 signals have PRN sequences that are about 6.2×10^{12} bits long (773 gigabytes) and take one week to transmit, even at a higher bit rate (or chip as it is called).

The civilian PRN is known as the C/A code (coarse/acquisition) and the military the P (precision) code.

There is also a Y and a more modern M code for military use. These have improved anti-spoofing and anti-jamming capabilities.

There are also modernised civilian services on later satellites, on the L2 frequency (called L2 CM and L2 CL) which offer improved navigational accuracy and other benefits. “Safety of life” signals are also transmitted on more recent satellites on the L5 band, along with PRN ranging codes.

Apart from enabling multiple signals on one frequency,

The role of GPS in timekeeping

Apart from its obvious role in navigation, GPS also plays a vital role in timekeeping via the very accurate atomic clocks each satellite has onboard.

GPS can provide accurate time to within nanoseconds, compared to the old radio signals that provided millisecond accuracy.

Many industries use GPS timekeeping services to:

- provide a time stamp on transactional records
- keep mobile phone networks synchronised
- keep power grids synchronised
- keep digital broadcast services operating correctly, allowing efficient utilisation of limited radio spectrum bandwidth
- allow scientific instruments distributed over wide geographic areas, eg, seismometer networks utilising a common time reference... and for many other uses.

Equation	Description
$\mu = 3.986008 \times 10^{14} \text{ m}^3 / \text{s}^2$	WGS 84 value of earth's universal gravitational parameter
$\dot{\Omega}_e = 7.292115167 \times 10^{-5} \text{ rad} / \text{s}$	WGS 84 value of earth's rotation rate
$a = (\sqrt{a})^2$	Semimajor axis
$t_{n+1} = t - t_{0e}$	Time from ephemeris reference epoch
$f_n = \tan^{-1} \left\{ \frac{\sqrt{(1-e^2)} \sin E_n / (1-e \cos E_n)}{(\cos E_n - e) / (1-e \cos E_n)} \right\}$	True anomaly
$E_n = \cos^{-1} \left(\frac{e + \cos f_n}{1 + e \cos f_n} \right)$	Eccentric anomaly from cosine
$\phi_n = f_n + \omega$	Argument of latitude
$\delta\mu_n = C_{\mu C} \cos 2\phi_n + C_{\mu S} \sin 2\phi_n$	Second-harmonic correction to argument of latitude
$\delta r_n = C_{rC} \cos 2\phi_n + C_{rS} \sin 2\phi_n$	Second-harmonic correction to radius
$\delta i_n = C_{iC} \cos 2\phi_n + C_{iS} \sin 2\phi_n$	Second-harmonic correction to inclination
$\mu_n = \phi_n + \delta\mu_n$	Corrected argument of latitude
$r_n = a(1 - e \cos E_n) + \delta r_n$	Corrected radius
$i_n = i_0 + \delta i_n + (IDOT)t_n$	Corrected inclination
$x'_n = r_n \cos \mu_n$	X coordinate in orbit plane
$y'_n = r_n \sin \mu_n$	Y coordinate in orbit plane
$\Omega_n = \Omega_0 + (\dot{\Omega} + \dot{\Omega}_e)t_n - \dot{\Omega}_e t_{0e}$	Corrected longitude of ascending node
$x_n = x'_n \cos \Omega_n - y'_n \sin \Omega_n$	ECEF X coordinate
$y_n = x'_n \sin \Omega_n + y'_n \cos \Omega_n$	ECEF Y coordinate
$z_n = x'_n \sin i_n$	ECEF Z coordinate

CDMA (code division multiple access), a spread spectrum technique, is used to achieve this. CDMA was previously used on some mobile phone networks.

A PRN code is part of the CDMA scheme and is used to identify the signal of interest. It is a carefully selected binary code and one of a set. The PRN codes are chosen so that no two are alike.

The PRNs are called Gold codes after the person who invented them, and have “bounded small cross-correlations within a set” which means that they have the most possible difference between them (see Fig.3 & 4).

The PRN codes are predetermined and stored in both the satellites and receivers. By knowing the PRN code ahead of time, a receiver can pick out one signal from many that are simultaneously being received.

The PRN code is broadcast continuously, and the navigation data (at a much lower bit rate) is superimposed on that. The transmitted signal has more bandwidth than required for the transmitted navigation data, to allow the PRN code to be incorporated.

One way of thinking about this is like a room full of people all speaking different languages at the same time. If you are only interested in receiving the message of the speaker of one particular language (the desired PRN code),

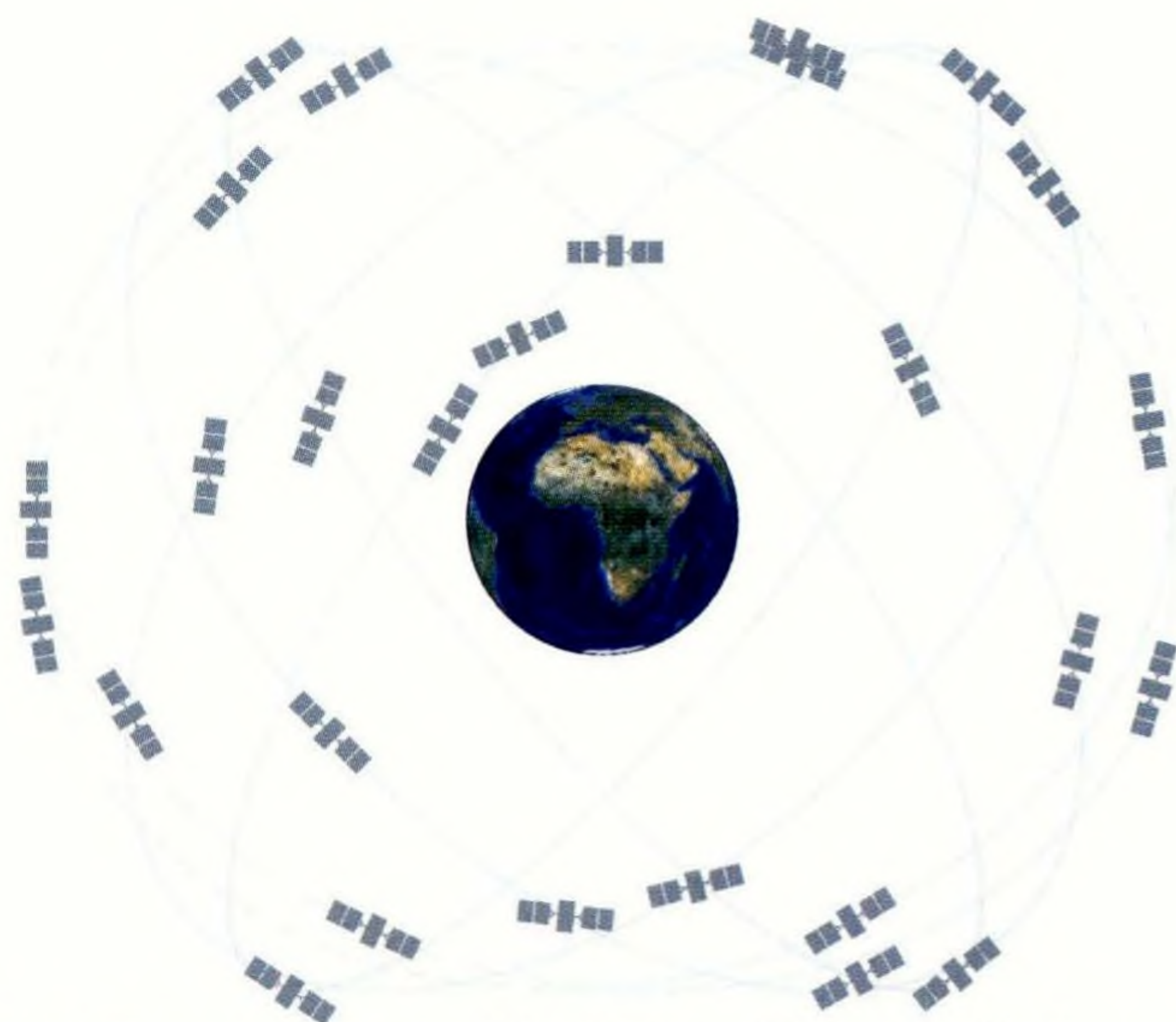
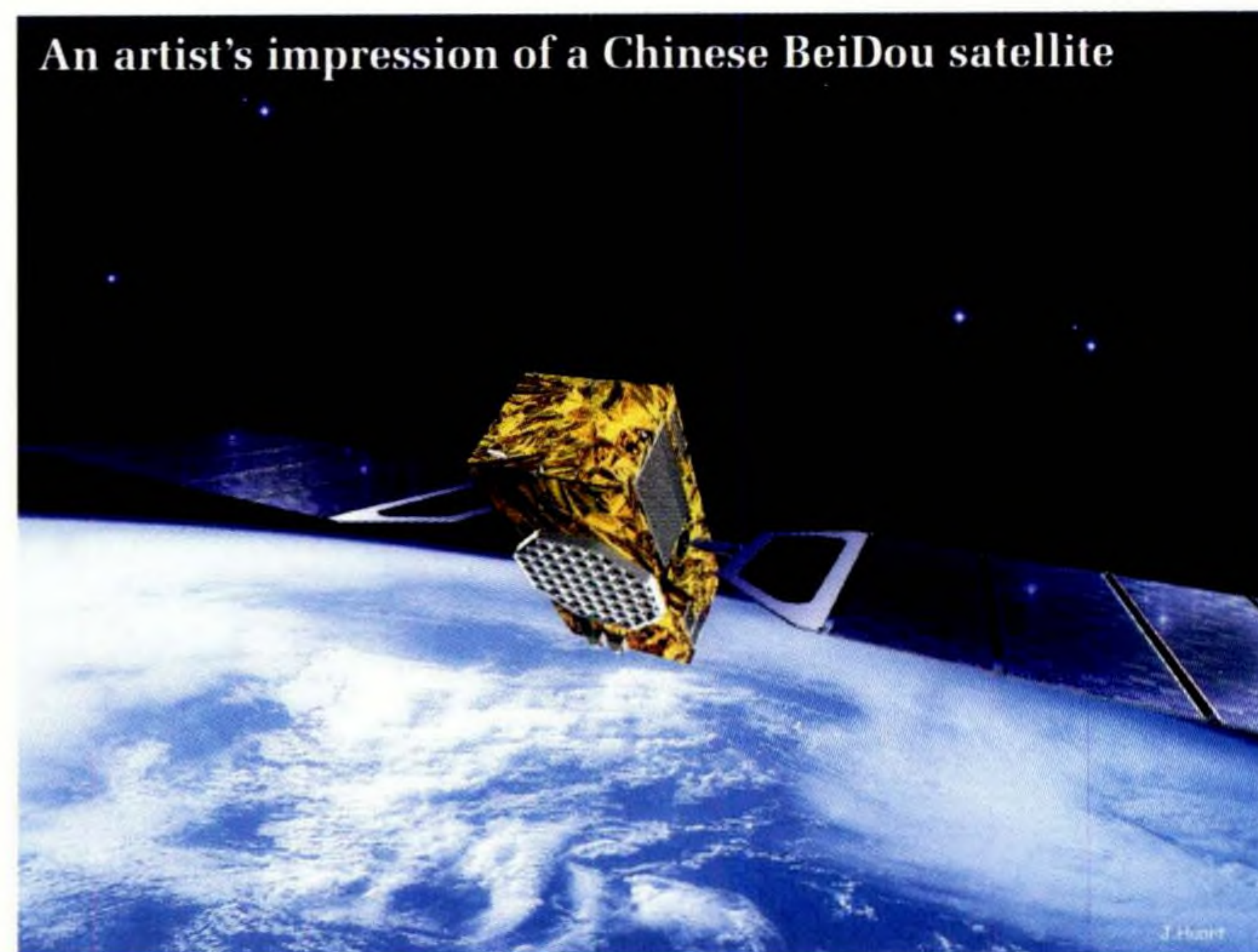


Fig.12: the arrangement of the GPS satellite constellation. The 24 satellites are in six equally spaced orbital planes and within each plane, there are four positions or “slots” occupied by satellites. This arrangement ensures that there are at least four satellites visible overhead at any point on the Earth’s surface at all times.

the CDMA technique allows for low transmission power and resistance to jamming and interference.

Applying time corrections

With the PRN code enabling the identification of individual satellites, and with knowledge of the PRN code expected at a particular time from that satellite in the receiver, it is possible to determine the offset between two matching segments of code and thus determine the approximate distance to a satellite. This distance is subject to corrections and thus called the pseudorange.



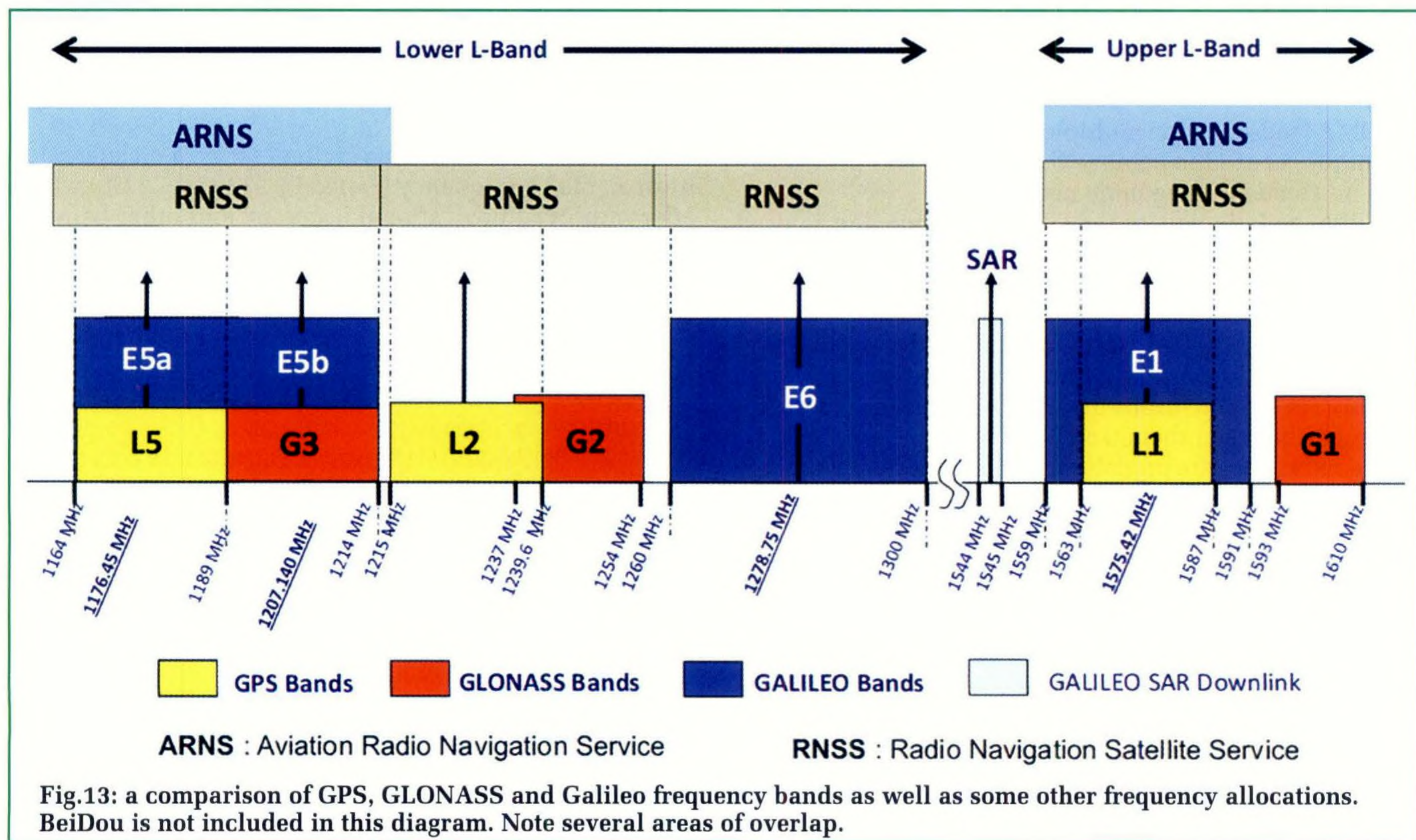
An artist's impression of a Chinese BeiDou satellite

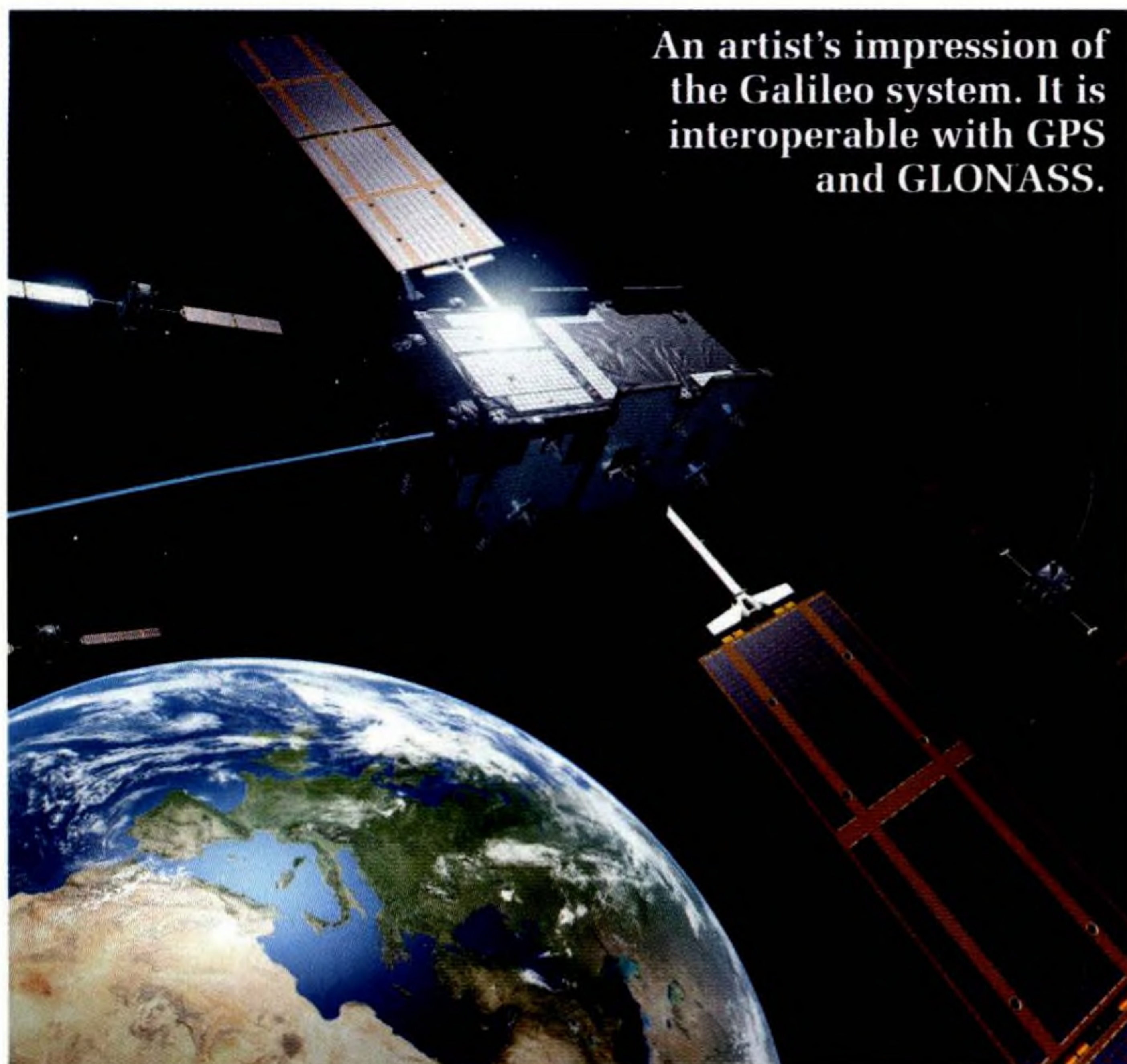
A typical receiver such as a hand-held unit, smartphone or in-vehicle navigation system does not have an atomic clock due to reasons of cost, size and power consumption. So the receiver is not precisely synchronised with the clock on the satellites, leading to uncertainty in the actual distance to the satellites.

The discrepancy between highly accurate clocks on the satellites and the less accurate clock at the receiver is resolved as follows.

The receiver gets signals from multiple satellites for a positional fix. The spheres representing the distance to three satellites will always intersect at two points (one of which is ignored), even if the clock receiver is wrong.

With a fourth satellite and a fourth sphere representing that satellite, there can only be one value of receiver time that satisfies the condition of the four spheres intersecting at one point.





An artist's impression of the Galileo system. It is interoperable with GPS and GLONASS.

The receiver adjusts its clock to that value, thus ensuring it is synchronised to the atomic clocks of the satellites. It must do this constantly due to the inherent inaccuracy in the receiver clock.

Further corrections

There are signal delays caused by the passage of signals through the ionosphere, leading to an error of about 83ns or 25m. Passage through the troposphere leads to an additional 7ns or 2m error.

Other effects taken into account either on the satellite or user equipment are the space vehicle clock error due to bias, drift, ageing and group delay (the time delay as a function of frequency for a signal to go through the electronics of the satellite).

A detailed flow chart for the process is shown in Fig.5. As can be seen, the process involves calculating the satellite clock bias, group delay, relativistic and other effects; accounting for delays due to the atmosphere and geometric effects and then correcting pseudorange to give the Estimated Range Deviation (ERD) due to these timing effects.

But there are still more calculations that need to be made



An artist's impression of a QZSS satellite in orbit.

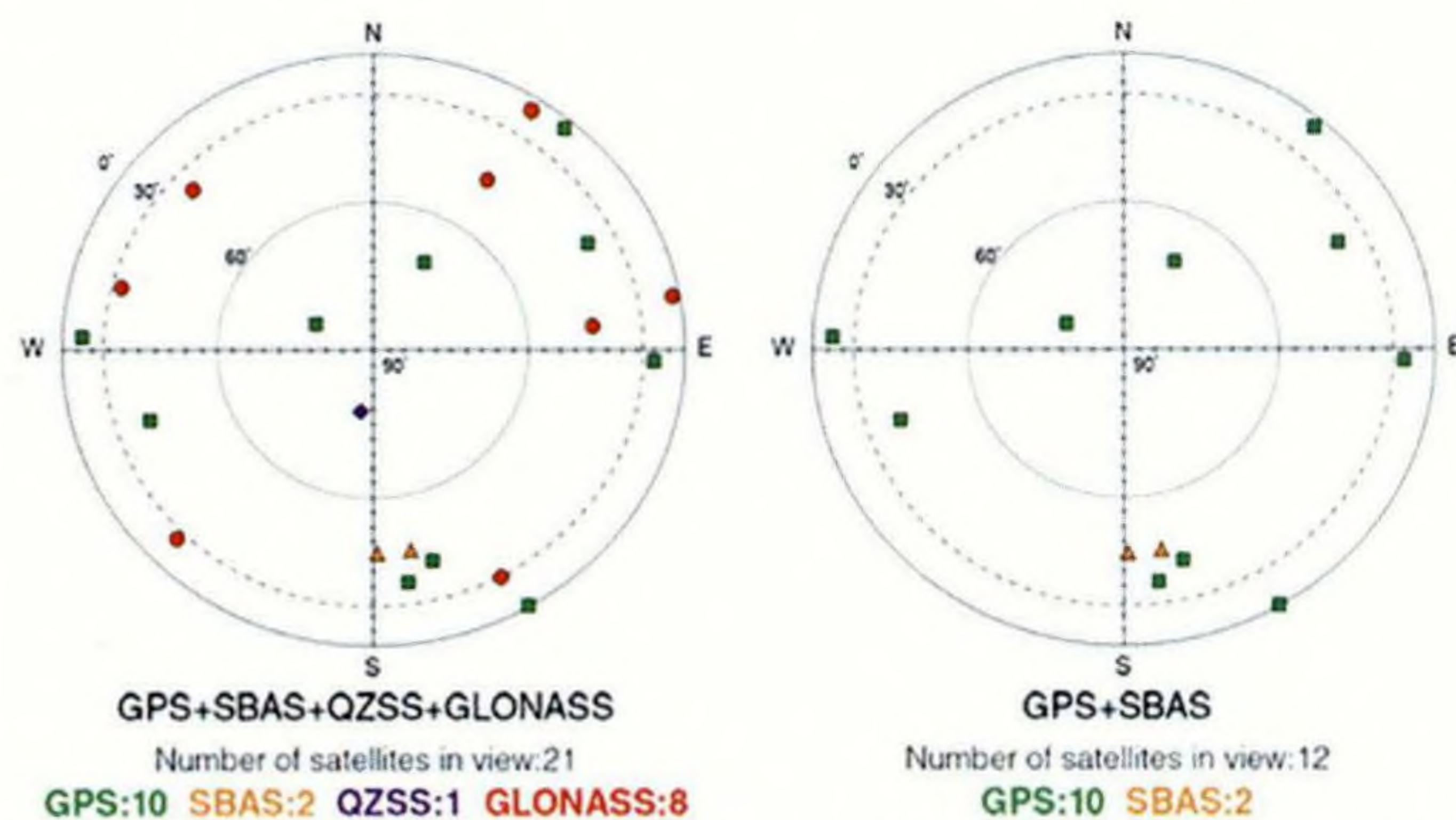


Fig.14: an example of satellites visible from Furuno's multi-GNSS receiver, compared to their GPS-only receiver. SBAS stands for Satellite Based Augmentation System.

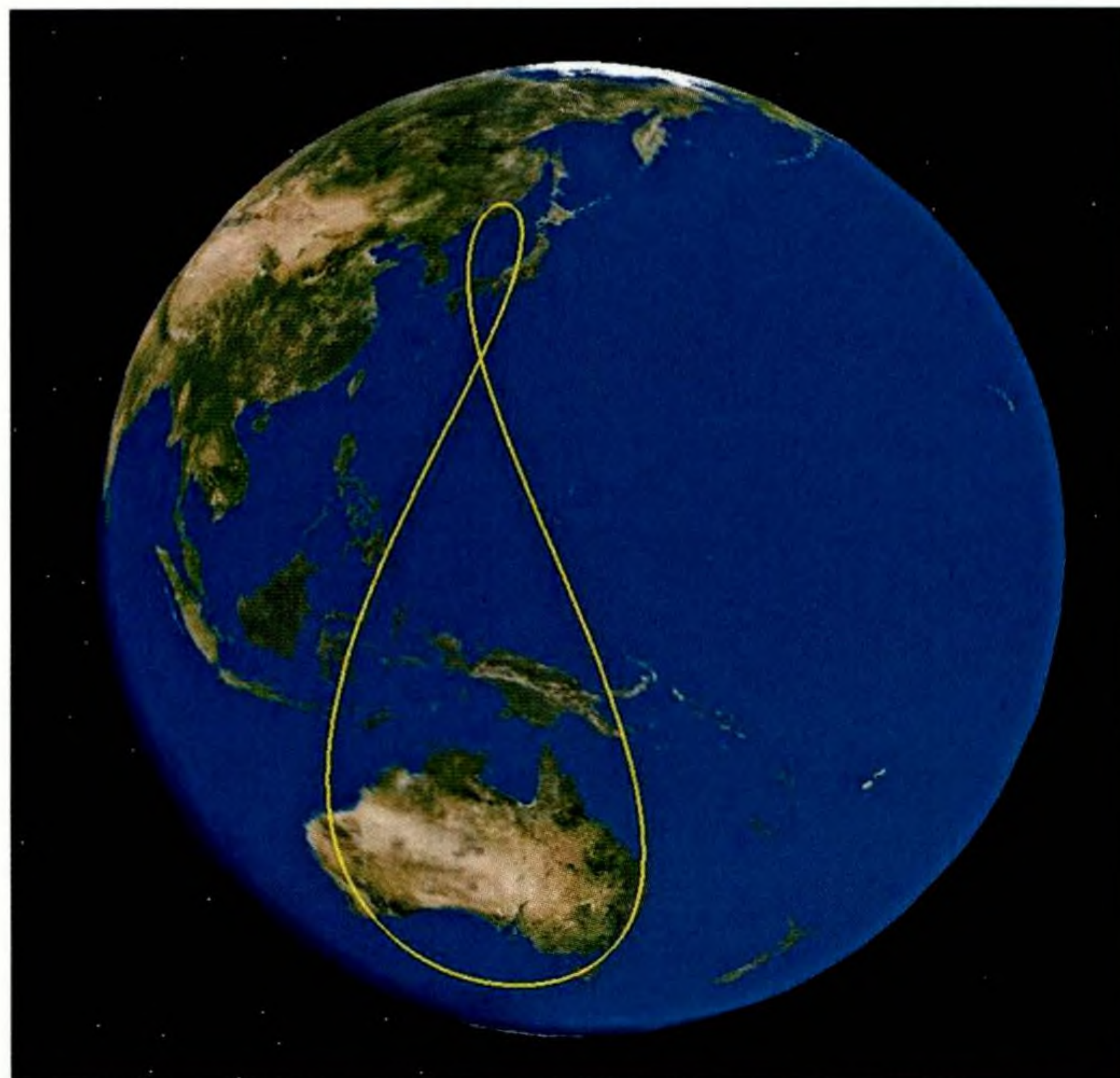
to determine the receiver position.

Establishing the position of the satellites

We now have explained how the pseudorange is obtained and how this is adjusted to get the exact propagation time of the radio signals from the satellite to the receiver. We still have not established the positions of the satellites from which we can determine a navigational fix, and that is the next task. All GNSS systems work similarly but we will look at GPS as an example.

A GPS satellite transmits both an almanac (Fig.6), which has general positional data for all satellites in the constellation, and an ephemeris (Fig.7), which is accurate orbital data specific to itself. Almanac data gives information about what satellites are in the area of view of the receiver, so that it doesn't have to spend extra time looking for satellites that are not visible.

Almanac data is considered valid for about 180 days, while ephemeris data is valid for only about four hours. Ephemeris data allows the exact location of a satellite to be established at a given time, and takes into account perturbations due to gravitational influences on the satellite from bodies such



A ground track of one of Japan's QZSS satellites, showing that most of Australia is included in its coverage area. This effectively adds four GPS-compatible satellites.

DIY GNSS Projects

GPS modules can be purchased from the SILICON CHIP ONLINE SHOP. If you can afford to wait, they're also available quite cheaply on line, for as little as \$5 delivered.

SILICON CHIP has published many projects using inexpensive GNSS modules for purposes such as clock and frequency references, or most recently as an accurate speedo and automatic vehicle audio volume controller (June 2019). See: www.siliconchip.com.au/project/gps

Here are online articles explaining how to interface these modules to Arduinos, Raspberry Pis and other similar platforms.

- To connect to a GNSS module to a PC, you need a USB/serial converter. Make sure it is the correct voltage for the module, usually 3.3V. Note that some converters won't work with Windows 10; the CP2102-based modules in our Online Shop are relatively trouble-free.
- A popular GNSS module brand is u-blox. They make the VK2828U7G5LF modules sold in our ONLINE SHOP (Cat SC3362). They have free evaluation software that allows you to see many aspects of GNSS operation with their modules.
- See: siliconchip.com.au/link/aavm (Windows software). A suitable Arduino library called TinyGPS++ is at: siliconchip.com.au/link/aavn
- See the videos titled "Playing with GPS: Ublox Neo-7M and U-Center" at: <http://siliconchip.com.au/link/aavo> and "10Hz U-blox binary GPS data in 66 lines of code (Arduino)" at siliconchip.com.au/link/aavp
- You can see the position of various GNSS or other satellites in the sky at: siliconchip.com.au/link/aavq
- Build a geocaching pendant as described at: siliconchip.com.au/link/aavr
- RTKLIB (www.rtklib.com) is an open-source program for high-precision GPS with low-cost devices. See also: <http://rtkexplorer.com/>

as the sun and moon.

Ephemeris data includes the standard six Keplerian elements, plus ten others, to take into account minor influences which affect the satellite's orbit.

The orbit of a satellite can be determined using the laws of physics plus minor deviations from theory due to unknown random forces, which are determined with ground-based radar, providing corrections incorporated into the ephemerides (the plural of ephemeris). Fig.8 shows the meanings of the ephemeris parameters, while Fig.9 shows the calculations involved.

The coordinate system

Once a GNSS system has established the receiver position, it still needs to be placed on a particular reference frame. The Earth is not a sphere but rather an "oblate spheroid" of 6,378,137m x 6,357,002m.

Various standard reference frames have been developed for navigation that correctly place coordinates on the Earth's true surface.

Typically, WGS84 (World Geodetic System) is used for



Fig.15: a screen grab of the GPSTest App for Android showing satellites visible on the phone's GNSS receiver. From top to bottom, the constellations are GPS, GLONASS, Galileo and BeiDou.

GPS for the so-called Earth-centred, Earth-fixed (ECEF) reference frame. In Australia, the standard reference frame for high precision work is the Geocentric Datum of Australia, GDA94 but as Australia drifts north due to tectonic plate movements, this is now out by 1.6m.

GDA2020 is under development; WGS84 still gives acceptable results for most users.

The atomic clocks

GNSS systems would not be possible without the use

Comparison of satellite navigation systems

	GPS	GLONASS	Galileo	BeiDou
Country	USA	Russia	EU	China
Number of satellites as of 18 June 2019	total: 32 31 operational 1 in maintenance	total: 27 24 operational 1 in commissioning 1 spare 1 in testing	total: 30 26 operational 4 to be launched (3 of which are spares)	total: 39 33 operational 6 non-operational 35 to be operational by 2020
Altitude	20180km	19130km	23222km	21150km for medium Earth orbit satellites (MEO)
Frequencies used	L1: 1575.42MHz L2: 1227.60MHz L3: 1381.05MHz L4: 1379.913MHz L5: 1176.45MHz (L1 and L2 are the primary frequencies, others are little used or experimental)	Modernised: L1: 1600.995MHz L2: 1248.06MHz L3: 1202.025MHz For future interoperability with other systems: L1: 1575.42MHz L3: 1207.14MHz L5: 1176.45MHz	E1: 1575.420MHz E6: 1278.750MHz E5: 1191.795MHz E5a: 1176.450MHz E5b: 1207.140MHz	B1I, B1Q: 1561.098MHz B1C, B1A: 1575.42MHz B2B, B2I, B2Q: 1207.14MHz B2a: 1176.45MHz B3I, B3Q, B3A: 1268.52MHz
Signal encoding	CDMA	FDMA but moving to CDMA	CDMA	CDMA
Orbital period	11h 58m (half a sidereal day)	11h 15m	14h 7m	For 27 satellites in MEO: 12h 37m
Orbital regime	6 planes in medium Earth orbit	3 planes separated by 120°, 8 satellites in each plane; satellite inclination 64.8°	27 operational satellites in 3 planes with 56° inclination to the equatorial plane	For 2020: 5 geostationary 3 inclined geosynchronous 27 Medium Earth orbit
Accuracy	300-5000mm	2.8-7.38m, next-generation GLONASS-K2 from 2019 is intended to reduce user range error to 300mm	1m public 10mm restricted	10m public (global) 5m Asia Pacific region 100mm restricted
First in service	First launch: 1978 Initial operational capability: December 1993. Fully operational: April 1995	Claimed fully operational in December 1995 but not globally available until the mid-2000s	Completion by end 2020 but operational now	Completion by end 2020 but some services available since December 2012

of extremely accurate atomic clocks. As mentioned above, radio signals travel 30cm in one nanosecond, so clock accuracy has to be of that order or better to obtain a good navigational fix.

GPS satellites have four onboard cesium and rubidium atomic clocks. These are kept in sync and are adjusted by even more accurate Earth-based atomic clocks. Typical accuracy of the clock on the latest GPS satellites is ± 4 nanoseconds, representing about ± 120 cm of range error.

We published an article in the February 2014 issue which explained how rubidium atomic clocks work (siliconchip.com.au/Article/6127).

GNSS receiver start-up

A receiver usually cannot get a position fix as soon as it is powered up. There are three distinct start-up situations which lead to differing power-on times before a fix can be made.

If the receiver is brand new or hasn't been used for a long time, that makes it a 'cold start'. The receiver doesn't know where it is, so it has to search for all possible satellites. After a satellite is acquired, it then has to download the almanac data for all satellites. This takes 12.5 minutes and gives it the approximate positions of the other satellites.

A 'warm start' is where the receiver already knows the time within 20 seconds and its position within 100km and

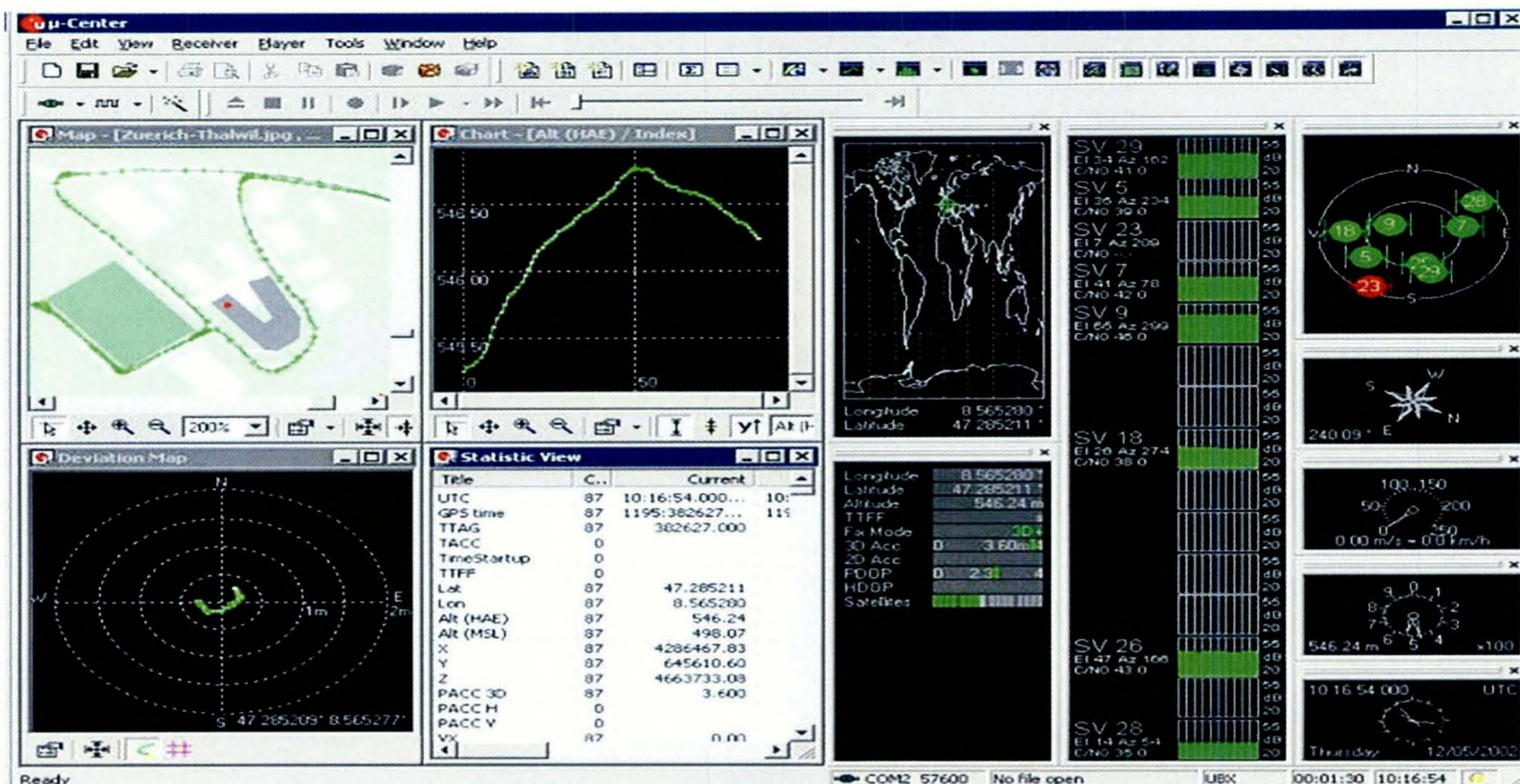


Fig.15: an example of the free u-center evaluation software for Windows, which allows inexpensive u-blox GNSS modules to be tested and configured. See siliconchip.com.au/link/aavm

has current almanac data.

It can then find the ephemeris data for at least four satellites, which is broadcast every 30 seconds, and then get a positional fix, usually within a minute.

A 'hot start' is where the receiver has current time, almanac, ephemeris, and position to allow rapid acquisition of new signals, usually within a few seconds. Vehicle GPS systems which can be "always on" may use this system.

Some GNSS systems used in Smartphones can sometimes start faster, because in addition to the GNSS location, they also use a database of WiFi network locations to help determine their location earlier than the GNSS signal would permit.

GPS and leap seconds

The global time standard is UTC or Coordinated Universal Time. Since the Earth's rotation rate varies naturally by a slight amount, every so often a leap second is added or removed to keep Universal Coordinated Time synchronised with the Earth's rotation.

The leap second is not implemented in GPS because of the navigational errors and confusion this would cause.

The difference between UTC and GPS time was zero when the GPS clock started on 1st January 1980, but is now 18 seconds.

The GPS Navigation Message broadcasts the difference between UTC and GPS time, so a receiver can show the correct UTC or local time.

Mapping errors

Finally, note that while a GNSS fix is generally extremely accurate, the maps used by navigation systems are not necessarily accurate.

There have been many mishaps due to people following incorrect maps, only to become stranded, or in some cases, driven over cliffs or off the end of piers! This is, of course, a problem of the maps and not the GNSS system itself.

To help ensure the most accurate possible and free maps the public can contribute to the production of open-source maps by joining the OpenStreetMap community (www.openstreetmap.org).

Some maps contribute to specialised interests such as four-wheel-driving, mountain biking, bush walking, etc, while others concentrate on regular street navigation.

Mixing and matching multiple GNSS systems

Many modern satnav receivers can decode GPS, Galileo, GLO-NASS and QZSS (Japan's regional system) signals. An increasing number of devices can also decode BeiDou.

Multi-GNSS receivers have improved performance due to the greater number of satellites in view, especially in urban canyons where the view of the sky is very limited.

You can see what systems your Android smartphone can receive with the free GPSTest App. Note that QZSS, which is visible in most of Australia with standard GPS receivers, effectively adds four more satellites to the constellation

Novel wearable GPS products

One of the more special GPS products we have seen is the GPS SmartSole, a GPS unit built into the sole of any shoe that connects to the mobile phone network. It can be used to track loved ones with memory disorders.

See: siliconchip.com.au/link/aavs
Their tracking services are available in the USA only.

They do not mention whether it is compatible with a shoe phone, so if you are hot on the heels of a KAOS agent, you had better do your own testing!