

451-337 Satellite Positioning and Geodesy Exam Solutions 2007

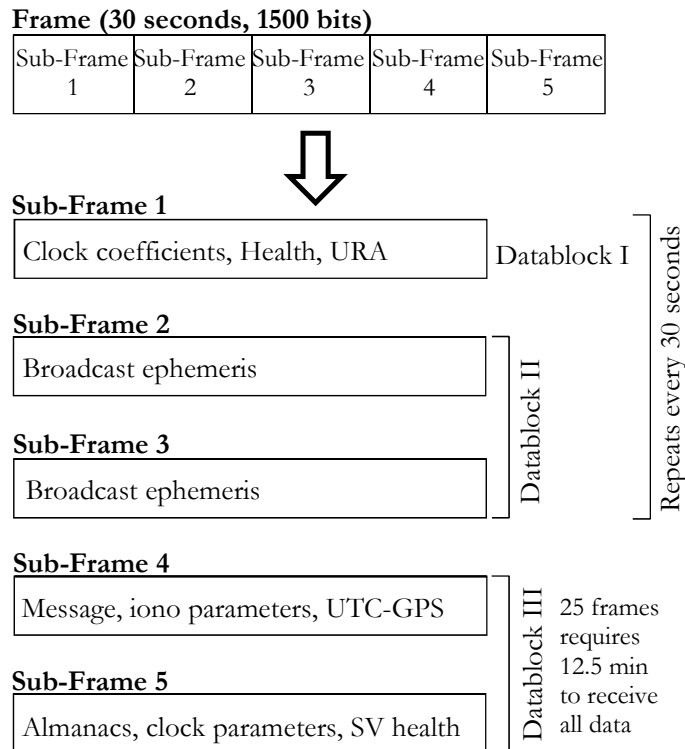
1. (a) The GPS *satellite message* provides a considerable amount of information to users of the system. Provide a complete description of the structure, contents and applications of the satellite message.

(8 marks)

The satellite message is a 50 bit per second data stream transmitted on the L1 and L2 carriers. It provides users with both satellite-specific and constellation wide information essential for the reliable computation of position determined by a GPS receiver. The satellite message is transmitted in a series of *frames*. Each frame is 30 seconds long (1500 bits) and is divided into five sub-frames each of 6 seconds duration (300 bits).

The first three sub-frames relate to the transmitting satellite and are repeated in each frame (every 30 seconds). They contain satellite-specific information such as clock coefficients, satellite health, URA and the 16 parameters of the broadcast ephemeris. Sub-frames 4 and 5 relate to other satellites in the constellation and contain ionospheric parameters, UTC-GPS time offset, almanac data, clock parameters and satellite health. 25 frames are required to transmit the full satellite message for all satellites in the constellation (25 satellites * 30 seconds = 12.5 minutes).

The following diagram illustrates the structure and contents of the satellite message:



- (b) Give complete definitions for the following:

- (i) Precise ephemeris
- (ii) Ambiguity resolution
- (iii) The geoid
- (iv) Geodetic coordinates
- (v) Conformal transformation

(12 marks)

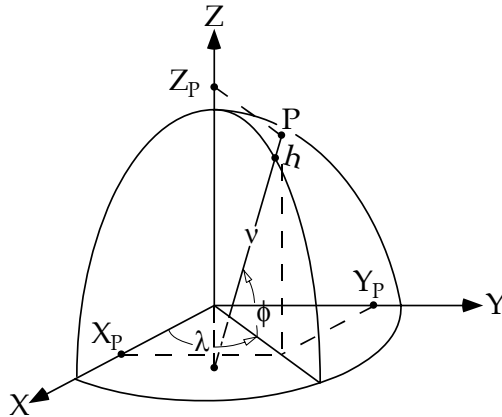
Precise ephemeris provides the actual (estimated) positions of GPS satellites in the form of ITRF (X,Y,Z) coordinates at 15 minute intervals. The precise ephemeris is obtained by post processing of satellite and atmospheric data collected by the IGS (International GNSS Service). There are over 200 globally distributed tracking stations that contribute to the IGS precise ephemeris computations. The disadvantage of the precise ephemeris is that it is not available for real-time positioning (except when using the predicted component of the ultra-rapid ephemeris), however, on the positive side, it does provide a substantially better estimate of satellite locations. Currently the IGS claims that satellite locations can be provided to about ± 0.05 m and satellite clock correction parameters to ± 0.1 nanoseconds, compared to ± 2 m and ± 7 nanoseconds respectively for the broadcast ephemeris. The precise ephemeris is made available free of charge over the internet in three forms:

- ☞ *Ultra-rapid* [predicted component] – available in real-time
- ☞ *Ultra-rapid* [observed component] – available with a 3 hour delay
- ☞ *Rapid* – available with a 17 hour delay
- ☞ *Final* – available with a 13 day delay

Ambiguity resolution refers to the process that allows the integer ambiguity associated with the GPS carrier phase observable to be determined computationally through the processing of GPS carrier phase data. While carrier phase measurements themselves are very precise, they are ambiguous in as much as the number of full cycles between the receiver and the satellite at the time of lock-on can never be measured. In order to turn the carrier phase measurement into a non-ambiguous range requires that this integer ambiguity be determined. This can only be done mathematically as part of the solution to determine receiver coordinates. The quality of the derived coordinates will be explicitly linked to the success or otherwise of the ambiguity resolution process. Reliable ambiguity resolution implies the determination of accurate and precise receiver satellite ranges and therefore optimum receiver coordinates. If uncertainty exists as to the integer ambiguities, the ranges will be inaccurate and the receiver coordinates will be compromised accordingly.

Geoid is an equipotential surface of the earth's gravity field. An equipotential surface is a surface of equal gravity potential to which the direction of the instantaneous gravity vector is at all points perpendicular. Because the gravity vector is always at right angles to an equipotential surface, this surface is – by definition – a *level* surface. Thus water will not flow across an equipotential surface. The geoid is coincident with the mean sea level surface of the open oceans, once the influences of tides, wind, currents and sea surface topography have been accounted for.

Geodetic coordinates are latitude, longitude and ellipsoidal height as illustrated in the following figure.



Latitude (ϕ) is an angle measured in the meridian plane from the equator (or the zero parallel) to the ellipsoid normal that passes through the point of interest (P in Figure 1). Latitude is reckoned positive north of the equator and negative to the south. Thus latitude will fall in the range $\pm 90^\circ$ or 90°S to 90°N .

Longitude (λ) is an angle measured in the equatorial plane from the zero (Greenwich) meridian to the meridian through P. Longitude is reckoned positive to the east of Greenwich and negative to the west. Thus longitude will fall in the range $\pm 180^\circ$ or 180°E to 180°W .

Ellipsoidal height (h) is the distance (in metres) reckoned along the ellipsoid normal through P between the surface of the ellipsoid and the point P.

Conformal transformation is a transformation of coordinates that maintains shape. In geodesy, the commonly used conformal transformation is the 3D seven parameter similarity transformation (otherwise known as the Bursa-Wolf or the Helmert transformation). The seven parameters include three translations (ΔX , ΔY , ΔZ), three rotations (R_x , R_y , R_z) and a scale factor (λ). A 2D equivalent has four parameters but is less commonly used. The conformal transformation model in geodesy is applied to the Cartesian coordinates (X, Y, Z) and moves them from one frame of reference (or datum) to another. The 3D conformal transformation is represented by the following equation:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} + (1 + \lambda) \begin{bmatrix} 1 & R_z & -R_y \\ -R_z & 1 & R_x \\ R_y & -R_x & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

2. One of the keys to optimising the precision and accuracy of GPS carrier phase positioning is to understand and manage the various sources of error. Describe the main errors that impact on GPS surveying and detail options for error mitigation. (20 marks)

The following table lists the major errors affecting GPS carrier phase observations. These errors have been classified according to their source: satellite, atmosphere and receiver.

Source	Error
Satellite	Orbit Clock
Atmosphere	Ionosphere Troposphere
Receiver	Clock Antenna phase centre Multipath Noise

Satellite orbit

For a variety of reasons, GPS satellites do not follow a perfectly defined path. Thus the real location of a satellite at the point of transmission of a signal may be different to its position as specified by the ephemeris data that is being used (precise or broadcast). This orbit error will impact on the computed range and thus on the estimated receiver coordinates.

The mitigation of satellite orbit error can be accomplished by doing relative positioning since this type of error has a high degree of spatial correlation. Alternatively, use of the precise ephemerides can be a very effective way of minimising the impact of satellite orbit errors. For example the IGS *Final* ephemeris has a nominal accuracy of ± 5 cm on satellite locations, compared to ± 2 metres using the broadcast ephemeris.

Satellite clock

While the atomic clocks carried by GPS satellites are very accurate, they are not perfect. These clocks are set to run fast to account for most of (99.6%) the relativistic effects and the MCS physically adjusts the satellite clocks to make sure they are within 1 millisecond of GPS time. Furthermore, the MCS computes and uploads clock correction parameters (three parameters of a second order polynomial) to further improve the accuracy of the GPS clocks to about ± 7 nanoseconds. However, this level of accuracy is still not sufficient for high precision carrier phase positioning. Thus the user must further deal with satellite clock errors.

Options for mitigating satellite clock errors include relative positioning, where the formation of a between-receiver single difference can totally eliminate the satellite clock error, or the use of precise IGS products, where the satellite clock error is modelled to an accuracy of ± 0.1 nanosecond.

Ionospheric delay

The ionosphere is that layer of the earth's atmosphere that extends from about 50 km to about 1000 km above the surface of the earth. The ionosphere is a dispersive medium since the amount of refraction depends on the wavelength of the signal. The ionospheric error is the largest single source of error impacting on GPS signals. The ionosphere is an electrically charged medium and ionospheric refraction is proportional to the number of free electrons, as measured by the Total Electron Content (TEC). Free electrons in the ionosphere are the result of a number of factors, but the most critical influence is solar radiation. A high incidence of solar radiation increases the number of free electrons and in turn increases the amount of refraction taking place. As a consequence ionospheric refraction varies with location (principally latitude) and time of day. It is worse in equatorial regions and at about the middle of the day (although the TEC maximum tends to lag a little behind the period of maximum solar radiation).

Ionospheric refraction affects the GPS codes and carriers in different ways. The carriers – being pure sinusoids – are affected by *phase velocity* and are accelerated in the ionosphere. The codes – being wave groups – are affected by *group velocity* and are thus slowed down when travelling through the ionosphere. An interesting feature of this diverse impact is that the advance caused to the carriers is equal in magnitude but opposite in sign to the delay of the codes. Thus if the phase advance causes the range to be shortened by k metres, this will be matched by an increase in the code pseudorange of the same amount.

Due to a tendency for the ionospheric range error to cancel in the horizontal plane, the ionospheric error mainly impacts on the determination of height.

There are two basic options for dealing with the impact of the ionosphere on carrier phase positioning.

Over short baselines (< 20 km), relative positioning will reduce the impact of the ionosphere due to the spatial correlation of this error source. This option is generally employed when doing single frequency GPS positioning.

Alternatively the vast majority of the ionospheric error can be modelled by utilising the fact that the ionosphere is a dispersive medium coupled with the fact that GPS satellites transmit on two frequencies. The dual frequency character of the system can be used to compute and largely eliminate the ionospheric error. This is the most comprehensive way of dealing with the ionosphere when processing carrier phase GPS data.

Tropospheric delay

The troposphere is that layer of the earth's atmosphere that lies between the surface of the earth and about 50 km above the earth's surface. It is sometimes called the *weather zone*. Unlike the ionosphere, the troposphere is electrically neutral and thus is a non-dispersive medium for GPS frequencies. The delay introduced by the troposphere affects the GPS codes and carriers in an identical way – both are slowed down. Thus the observation equations for code and carrier phase observables contain the same tropospheric delay term.

The tropospheric error is generally broken down into two components – the *dry* component (which represents about 90% of the error) and the *wet* component (making up the remaining 10%). The dry component is stable and relatively easy to account for using standard tropospheric models such as Hopfield and Saastamoinen. The wet component, though small, is highly variable and very difficult to model satisfactorily. Ignoring the tropospheric error tends to impact particularly on height and imposes a significant limitation on the usefulness of GPS for high precision height surveys such as subsidence monitoring.

Residual tropospheric errors (after dry modelling) can be minimised using relative positioning over short baselines, but the spatial correlation in the troposphere often only applies over quite short baseline lengths (depending on atmospheric variations).

An alternative is to use mathematical modeling to solve for residual tropospheric delay as part of the solution process – though this can be quite a complex process mathematically and computationally.

Receiver clock offset

The clocks used in GPS receivers are usually much lower quality than the satellite clocks. As a consequence, the receiver clock error can be very significant and must be dealt with in all forms of GPS positioning in order to remove this influence from the measured ranges.

The receiver clock error can be determined as part of the navigation solution on an epoch-by-epoch basis and this is adequate for many lower accuracy applications. However for high precision positioning, by taking between-satellite single differences when forming the double difference carrier phase observable, the receiver clock error can be totally eliminated.

Antenna phase offset and variation

When doing GPS positioning, great care is taken to locate the physical centre of the GPS antenna in relation to the survey mark or the point for which coordinates are required. The assumption behind this procedure is that the physical centre of the antenna is the place where the data from the satellites is received and measurements are made. This assumption is however incorrect. The reality is that the antenna phase centre is the point of measurement and the phase centre and the physical centre do not (necessarily) coincide. To account for this non-coincidence, the spatial relationship between the physical centre and the phase centre must be known. However this relationship is not constant, in fact it is usual to break it down into two components, a constant *offset* and the spatially changing *variation* component. The phase centre offset is a constant displacement (in 3D) between the physical centre and the phase centre. But the phase centre also moves as a function of satellite geometry (particularly satellite elevation) and this is the variable component.

Dealing with antenna offset and variation can be done in two ways. If both antennas are identical, parallel antenna orientation will result in a cancellation of the offset and variation errors through the process of measurement differencing. It is for this reason that GPS antennae are traditionally oriented to the north. The second option, which must be used in the case of using dissimilar antennae, is to employ antenna calibration models to correct the raw measurements for the influence of the error. The National

Geodetic Survey (NGS) in the US has a web-site devoted to this problem and supplies calibration models for a very wide range of GPS antennae. With the advent of CORS networks, the need to accommodate for the use of different antennae is becoming more common.

Receiver multipath

Multipath errors occur when GPS signals arrive at an antenna via an indirect path, having been reflected off a nearby surface such as the ground, a wall or some other reflective surface. The direct and indirect signals cause interference at the antenna and thereby corrupt the carrier phase measurements.

Multipath is not a spatially correlated error source and can, as a result be very difficult to deal with. Options include longer observing sessions to average out the multipath error, antenna design to mitigate the reception of reflected signals (e.g. choke ring antennae); and software mitigation strategies in the signal processing realm.

Receiver noise

There is nothing that can be done to change the noise characteristics of a particular receiver. Of course, increased session lengths will tend to average out the affects of random noise, but no modeling or observational strategies can be used to minimise noise affects.

3. (a) Measurement differencing is commonly used in the processing of GPS carrier phase observations. Describe the advantages and disadvantages of measurement differencing.

(5 marks)

Measurement differencing endeavours to take advantage of the fact that many of the errors that affect the collection of GPS carrier phase data are strongly spatially correlated. For example, consider the following simplified form of the GPS carrier phase observation equation.

$$\Phi_i^p = cdt^p - cdT_i + R_i^p - I_i^p + T_i^p - \lambda N_i^p + n_i^p$$

Where (i) represents the receiver and (p) the satellite. dt and dT are the satellite and receiver clock errors respectively, I and T represent the influence of the ionosphere and troposphere. N is the integer ambiguity and n is the combined receiver noise and multipath.

By combining observations between receivers and between satellites, many of the spatially correlated errors can either be eliminated or minimised. The following briefly summarises the main advantages and disadvantages of measurement differencing (presuming we are forming the double difference between two satellites and two receivers):

Advantages :

- ☞ Receiver clock errors are eliminated
- ☞ Satellite clock errors are eliminated
- ☞ Ionospheric errors are reduced (minimised)
- ☞ Tropospheric errors are reduced (minimised)

Disadvantages :

- ☞ Noise propagates (increases) because it is a random quantity
- ☞ Four individual carrier phase observations are needed for one DD (thus redundancy is reduced or more data must be collected)

(b) The observation equation for GPS carrier phase measurements is given below:

$$\Phi = c(dt - dT) + R - d_{\text{ion}} + d_{\text{trop}} - \lambda N(t_0) + n$$

- (i) Develop an observation equation for the triple difference observable.
- (ii) Describe the main advantages of the triple difference observable in processing GPS carrier phase data.

(15 marks)

(i) Note, we begin by introducing a slight simplification of the notation and have two carrier phase observation equations between two receivers (i) and (j) and a single satellite (p):

$$\Phi_i^p = cdt^p - cdT_i + R_i^p - I_i^p + T_i^p - \lambda N_i^p + n_i^p$$

$$\Phi_j^p = cdt^p - cdT_j + R_j^p - I_j^p + T_j^p - \lambda N_j^p + n_j^p$$

Forming the *between receiver* single difference gives :

$$\Delta\Phi_{ij}^p = -c\Delta dT_{ij} + \Delta R_{ij}^p - \Delta I_{ij}^p + \Delta T_{ij}^p - \lambda\Delta N_{ij}^p + \Delta n_{ij}^p$$

If the same two receivers track a second satellite (q), a second *between receiver* single difference can be formed as follows :

$$\Delta\Phi_{ij}^q = -c\Delta dT_{ij} + \Delta R_{ij}^q - \Delta I_{ij}^q + \Delta T_{ij}^q - \lambda\Delta N_{ij}^q + \Delta n_{ij}^q$$

The *double difference* is the difference between two single differences. Thus if we take the above between-receiver single differences, we can form a double difference as follows (note the same could be done with two between-satellite single differences) :

$$\Delta\nabla\Phi_{ij}^{pq} = \Delta\nabla R_{ij}^{pq} - \Delta\nabla I_{ij}^{pq} + \Delta\nabla T_{ij}^{pq} - \lambda\Delta\nabla N_{ij}^{pq} + \Delta\nabla n_{ij}^{pq}$$

Note that an identical equation could be written for the double difference computed for the next epoch of observation. Taking the difference between these two double differences gives the triple difference, which effectively eliminates the integer ambiguity term – presuming there have been no cycle slips. The observation equation for the *triple difference* observable thus appears as follows:

$$\delta\Delta\nabla\Phi_{ij}^{pq} = \delta\Delta\nabla R_{ij}^{pq} - \delta\Delta\nabla I_{ij}^{pq} + \delta\Delta\nabla T_{ij}^{pq} + \delta\Delta\nabla n_{ij}^{pq}$$

(ii) The main benefit of forming the triple difference is that when a triple difference solution is computed, the triple difference residuals can be analysed to find and fix cycle slips in the raw observations. Thus the triple difference allows the data to be “cleaned” prior to computing the more accurate double difference solution to resolve the integer ambiguities and determine precise receiver coordinates.

4. (a) GDA94 is the Geocentric Datum of Australia. Using GDA94 as an example, describe the processes of datum definition and datum realisation.

(10 marks)

Definition of GDA94

The definition of GDA94 was premised on the desire in Australia to have a geocentric datum that would be *a priori* compatible with satellite positioning systems such as GPS. In general terms, datum definition consists of the following four steps:

- ☞ *Size* of the ellipsoid
- ☞ *Shape* of the ellipsoid
- ☞ *Location* of the ellipsoid
- ☞ *Orientation* of the ellipsoid

With specific reference to GDA94, the chosen ellipsoid was GRS80 with internationally accepted values for the semi-major axis (a) and the flattening ($1/f$) thus establishing *size* and *shape*. With regard to location, the new datum was to be geocentric, so that the centre of the ellipsoid, and thus the origin of the accompanying Cartesian axes, would be coincident with the centre of mass of the earth. The orientation of the ellipsoid was chosen so that the XY-plane would be coincident with the equator, the XZ-plane coincident with the Greenwich meridian (reference meridian for longitude) and the Z-axis passing through the Conventional International Origin (CIO).

Realisation of GDA94

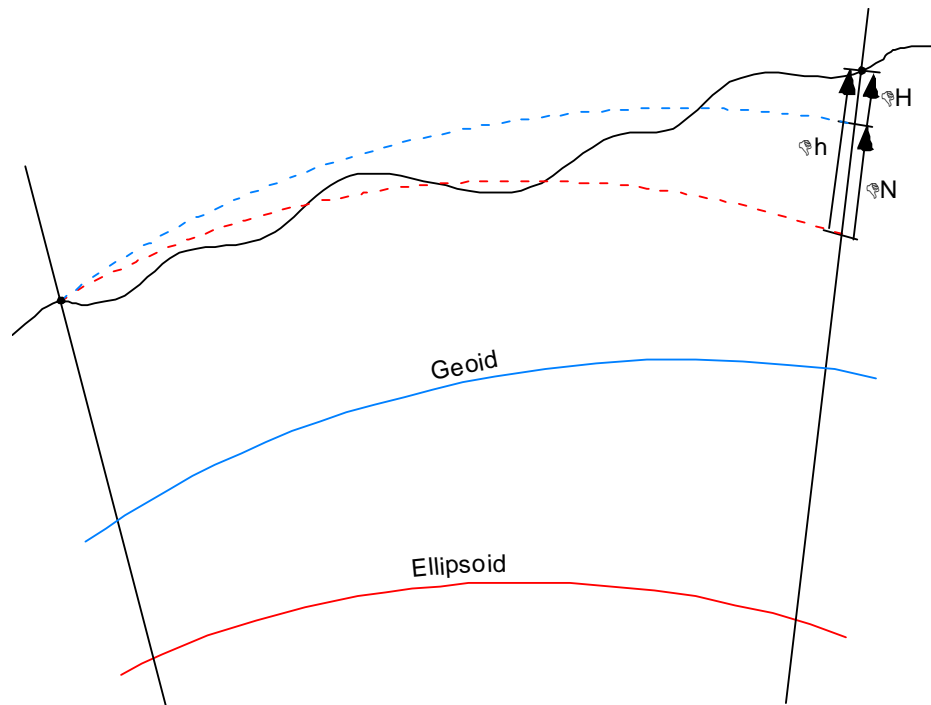
Again there are four stages in the realisation of a geodetic datum. Each stage will be discussed with specific reference to the realisation of GDA94.

- ☞ *Monumentation* – This stage involved the establishment of ground marks for which GDA94 coordinates would ultimately be determined. In the first instance, the monumentation stage involved the construction of the Australian Fiducial Network (AFN) sites at about 8 separate locations around Australia. Each site consisted of a substantial concrete pillar and a permanently operating GPS receiver. Below the AFN was the Australian National Network (ANN) which consisted of points with an average spacing of about 500 km and which were occupied by GPS receivers and connected by measurement to the AFN.
- ☞ *Observation* – This stage involved firstly the linking of the AFN to stations in the IGS network, for which ITRF92 coordinates were known. This measurement stage allowed direct connection between the Australian continent and an international geocentric frame of reference. A subsequent campaign of GPS observations allowed the ANN sites to be connected to the AFN.
- ☞ *Computation* – This stage involved processing of the GPS data and subsequent adjustment of the observed network to establish ITRF92 coordinates at the AFN and ANN sites. The coordinates were determined as at the epoch of January 1994. Thus GDA94 is, in effect, a realisation of ITRF92@1994.0 in Australia. It should be noted at this stage that GDA94 coordinates were subsequently propagated down to lower orders of the geodetic and survey control networks in Australia by readjustment of existing observations within a framework of GDA94 control.
- ☞ *Publication* – This final stage involved the public release of GDA94 coordinates for the various layers of survey control from AFN, through ANN and down through the hierarchy of control networks. Publication is the responsibility of federal and state survey and mapping agencies.

- (b) GPS is not able to directly determine orthometric heights. Elaborate on the reasons why this is so and describe how this so-called *GPS heighting problem* can be resolved. (10 marks)

The baseline that results from two simultaneously operating GPS receivers is expressed in cartesian components (ΔX , ΔY , ΔZ) relative to the WGS84 reference frame. When that baseline is converted into geodetic coordinate differences, it is expressed as ($\Delta\phi$, $\Delta\lambda$, Δh), where Δh is the ellipsoidal height difference between the end points. The ellipsoidal height difference is not a useful measure of “height” in conventional terms because ellipsoidal heights (and ellipsoidal height differences) do not relate to earth’s gravity field and therefore cannot be directly integrated into an existing height datum based on orthometric heights (such as AHD) nor can they be used to determine directions and rates of fluid flow. National height datums like the AHD are based on the geoid rather than the ellipsoid as a reference surface since the geoid is an equipotential surface of the earth’s gravity field and heights related to the geoid (denoted by H) can be reliably used to determine directions and rates of fluid flow. This dichotomy between ellipsoidal heights and orthometric heights lies at the core of the so-called GPS heighting problem and is illustrated in the following diagram.

In the figure, Δh the ellipsoidal height difference between two points as determined by GPS, is (or at least can be) different to the orthometric height difference ΔH determined by levelling because of the non-parallelism between the geoid and ellipsoid. The solution to the problem – that is to extract orthometric height differences from GPS – requires a knowledge of the relative geoid undulation (ΔN).



Solution of the GPS heighting problem is only possible by modelling the behaviour of the geoid with respect to the ellipsoid. Various options – some of them very complex

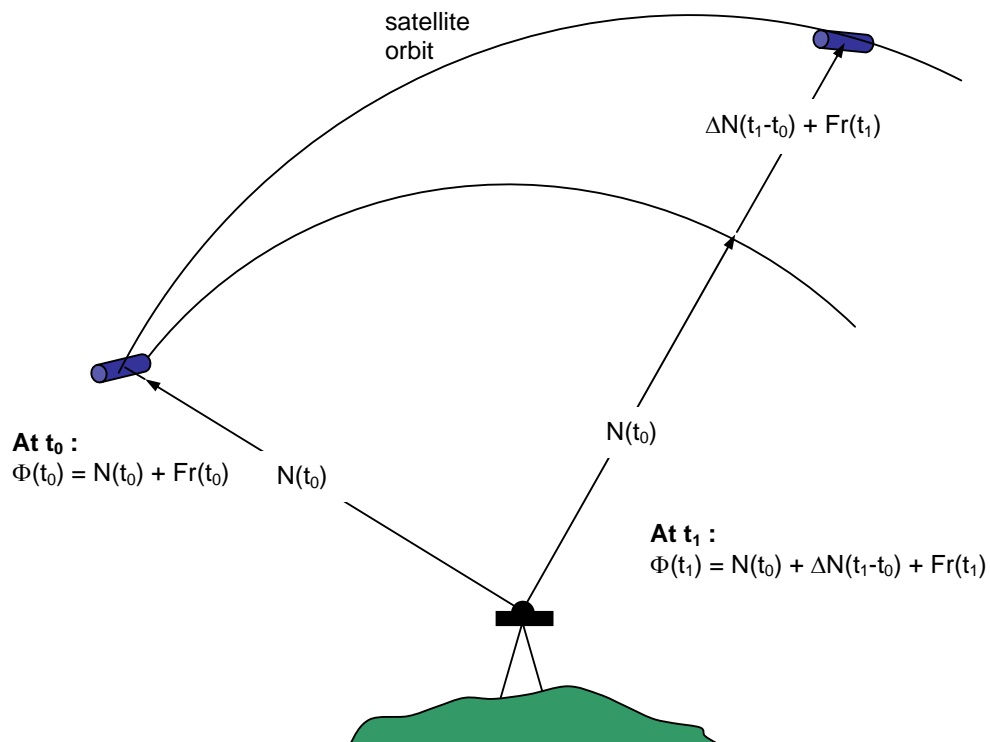
– exist to solve this problem. A common and quite effective approach in Australia, is to use the national geoid solution known as AusGeoid98, which is published by Geoscience Australia. AusGeoid98 provides geoid undulations at nominated points to a stated absolute level of accuracy of about ± 0.5 m, but relative geoid undulations are much more dependable at about ± 0.3 ppm. Whether from AusGeoid98 or some other approach – such as planar geoid modelling – once the relative geoid undulations are known, ellipsoidal heights from GPS can be converted into orthometric height differences and the third dimension from GPS relative positioning is thus made useful.

5. (a) Explain and illustrate what a GPS receiver actually measures when tracking the L1 and L2 carrier phase signals.

(7 marks)

The following diagram illustrates what happens when a GPS receiver tracks either the L1 or the L2 carrier wave transmitted by a GPS satellite. In the diagram, the epoch t_0 represents the initial epoch when the receiver locks onto the signal and commences to measure. At this epoch, the measurement is just the fractional component of the incoming sine wave ($Fr(t_0)$). There is no capacity within the receiver to measure the number of full cycles between the satellite and the receiver at this time. This unknown quantity is known as the *integer ambiguity* and is denoted by $N(t_0)$.

As the satellite passes overhead, it follows the orbit as shown, but the integer ambiguity remains constant so long as lock is maintained on the satellite signal. The constant integer ambiguity is illustrated in the diagram as a circular arc of radius $N(t_0)$ centred on the receiver. By maintaining continuous lock on the satellite signal, the receiver is able to keep count of the change in the number of full cycles so that at the next measurement epoch (t_1) the receiver is able to measure the fractional component of the in-coming wave ($Fr(t_1)$) and add this to the change in the number of full cycles which the receiver has counted ($\Delta N(t_1-t_0)$). Thus the receiver effectively measures the *change in range* to the satellite, leaving the initial integer ambiguity unmeasured and having to be resolved by mathematical techniques.



(b) Describe the GPS Standard Positioning Service (SPS).

(7 marks)

The GPS Standard Positioning Service is based on the transmission and reception of the C/A-code by civilian and other non-authorized users. The C/A-code is an unencrypted pseudo-random noise (PRN) binary stream, with a nominal wavelength of about 300 metres. Each satellite transmits a unique C/A-code allowing for satellite identification and (for all practical purposes) unambiguous range measurements between the receiver and the satellite. When the receiver is able to track at least four satellites simultaneously, it is possible to compute the 3D location of the receiver in the frame of reference of the satellites (WGS84) and the offset of the local receiver clock from GPS system time. The accuracy of this mode of positioning is about ± 10 metres or better (particularly horizontally).

(c) What changes are scheduled to be made as part of the GPS modernisation program and what benefits will these bring to users?

(6 marks)

The main features of GPS modernisation program include :

- ☞ Setting Selective Availability (SA) to zero
- ☞ New civilian code to be carried on L2, to be known as L2C
- ☞ New military code to be carried on L1 and L2, to be known as M-code
- ☞ New civilian frequency, to be known as L5
- ☞ Enhancements to the Control Segment (not specified)

The benefits of GPS modernisation include :

- ☞ Improved accuracy for all applications
- ☞ Improved integrity for critical users (e.g. civil aviation, automated aircraft landing systems)
- ☞ Improved signal availability
- ☞ Improved positioning integrity
- ☞ Faster and more dependable ambiguity resolution
- ☞ Better ionospheric modelling