

# Time and frequency distribution using satellites

**Judah Levine**

Time and Frequency Division and JILA, National Institute of Standards and Technology, The University of Colorado, Boulder, Colorado, USA

Received 11 April 2002, in final form 22 May 2002

Published 5 July 2002

Online at [stacks.iop.org/RoPP/65/1119](http://stacks.iop.org/RoPP/65/1119)

## Abstract

I will discuss how time and frequency information can be distributed using satellites. I will focus on using the signals transmitted by the US global positioning system satellites, but I will also discuss other satellite-based systems such as the Russian GLONASS system, the proposed European Galileo System and two-way satellite time transfer, which uses active ground stations that communicate with each other through an active satellite.

## 1. Introduction

In this paper, I will discuss how time and frequency information can be distributed using satellites. I assume that the reader is familiar with the general definitions of the International System of Units (SI) and the SI units for time and frequency. I also assume a familiarity with the general principles of operation of atomic frequency standards and with the statistical machinery (such as the Allan variance) that is used to characterize the noise processes in clocks and in a transmission medium. For an introduction to these subjects, see [1, 2]. I will focus on using the signals transmitted by the US global positioning system (GPS) satellites, since they are very widely used for time and frequency distribution, but I will also discuss other satellite-based systems in somewhat less detail. In particular, I will mention the Russian GLONASS system and the proposed European Galileo system, both of which are similar to GPS. I will also briefly discuss two-way satellite time transfer, which uses active ground stations that communicate with each other through an active satellite, whose transponders re-transmit any received signal (usually at a different frequency). This method, in which all sites are active, is fundamentally different from the other systems in which only the satellites do the transmitting while the ground stations are passive receivers.

In addition, there are a number of other satellite systems which I will not discuss. These include operational systems, such as GOES, which transmits time signals from weather satellites operated by the US National Oceanic and Atmospheric Administration (NOAA), and systems proposed for the future, such as the Atomic Clock Ensemble in Space (ACES), which will be based on a micro-gravity clock to be flown on the International Space Station. The accuracy of the received GOES signals is significantly poorer than any of the systems mentioned in the previous paragraph, and it is not clear at this time whether the cost of an ACES ground station will be low enough to make it generally affordable.

Satellite-based systems are not the only way of distributing time and frequency information, but they have two advantages that make them the systems of choice for many applications. The first is that a signal broadcast from a satellite can be received over a wide area, so that satellite-based systems can support time and frequency distribution over a large region much more easily than can be done by any purely terrestrial system. The second is that the path delay between a satellite-based transmitter and a ground-based receiver is usually more stable and can be more accurately modelled than the corresponding delay between the stations of a purely ground-based system. This is an important advantage, because the accuracy of any distribution method is almost always limited by the uncertainty in whatever method is used to compensate for the time it takes the signal to travel from the source to the destination.

Methods for path-delay compensation fall into two broad categories, depending on whether the signals travel in only one direction along the path, as is true for GPS, GLONASS and Galileo, or in both directions, as is true for two-way transmissions using communications satellites. If the signals travel only one way from the source to the receiver, then the path delay enters the uncertainty budget for the process in first order. Since the signal can travel no faster than the speed of light, the minimum travel time is  $3.3 \mu\text{s km}^{-1}$  of path. The path delay in a real system is generally larger than this value, since the signal propagates through a medium rather than a vacuum. This excess path delay is characterized by the refractivity of the transmission medium—the difference between the actual index of refraction and the value that it would have in vacuum, which is 1. The refractivity can vary due to changes in ambient conditions (such as temperature) and other factors. As we will see below, a typical value for the atmospheric refractivity is  $3 \times 10^{-4}$ , so that the presence of the atmosphere increases the delay by about  $1 \text{ ns km}^{-1}$  over what the path delay would have been if the signal had travelled over a vacuum

path of the same length. The refractivity depends on atmospheric pressure and temperature, and therefore can have significant variations in space and in time.

Instead of transmitting a signal between a single transmitter and a single receiver, it is also possible to transmit a signal from a single transmitter to two (or more) receivers simultaneously. Each receiver measures the time difference between its clock and the time as received from the transmitter, and these measurements are transmitted to an analysis centre where they are subtracted. If the receivers are approximately equidistant from the transmitter, then the path delays are also roughly equal, and the clocks at the receiving stations can be synchronized with an uncertainty that no longer depends in first order on the two path delays back to the single transmitter. Instead, the uncertainty is a function of the difference between the delays along the different paths, and both the magnitude of this difference and its fluctuations can be much smaller and easier to estimate than the full path delays themselves. This method, which is called ‘common view’, is very widely used both in satellite-based and terrestrial distribution systems.

In practice, the common-view paths to the receivers are never exactly equal, so that some ancillary corrections are necessary. The usefulness of this method depends on the equality of the time delays along the paths, on the correlation between the fluctuations in these delays and on how well any residual differences can be modelled or estimated. The method is very useful for satellite-based systems, because the fluctuations in the path delays to the receivers are often highly correlated so that these contributions cancel in the common-view difference. Although the method has also been applied to ground-based transmitters, it is often less useful in this situation because this correlation is often not nearly so high.

The common-view method has a number of other advantages. Since an offset in the time of the satellite clock tends to contribute equally to the measurements at all of the receiving stations, it cancels in the differences. This cancellation is not limited to static time offsets. Fluctuations in the time of the satellite clock are also cancelled or very strongly attenuated by the common-view subtraction process. This advantage of common view was especially important in the past, when the clocks on the GPS satellites were intentionally dithered to degrade their performance for non-authorized users. (The common-view method is not limited to time transfer. It is also used in geodetic applications, where it is called differential GPS.)

When the signal is transmitted in both directions along the path, the usual strategy is to assume that the delays are the same in both directions. There is no need to model the path delay at all with this assumption—the one-way delay is simply one-half of the measured round-trip value. Likewise, effects due to refractivity, fluctuations in the path delay and similar effects are assumed to affect both directions in the same way so that the assumption of symmetry is not affected. This assumption is most easily satisfied in the full duplex configuration, in which signals are transmitted in both directions simultaneously. If the system supports transmissions in only one direction at a time (half-duplex), then the transmission direction must be reversed periodically to measure the round-trip delay. These reversals must be made rapidly enough so that the path delay can be considered to be constant during the time interval required to make the measurement in both directions.

Although the assumption that the delay is the same in both directions is simple, it is not always correct. The largest contribution to the asymmetry usually comes from the hardware at the stations. For example, it is difficult to match the temperature sensitivities of the transmitter and the receiver at the stations, so that even if the delays in both directions are equal initially, they do not remain equal as the ambient temperature changes. Furthermore, since the satellite contains an active transponder rather than just a passive reflector, any asymmetry in the satellite hardware also contributes to the uncertainty budget. (An uncertainty in the delay through the station hardware or a variation in this delay caused by changes in the ambient

temperature are not unique to the two-way method; GPS receivers obviously have the same sort of problems, and these delay changes are not attenuated by a common-view subtraction.)

In the following discussion, I will start with the general considerations that are common to all of the one-way satellite systems: GPS, GLONASS and Galileo. I will then discuss the technical differences between these systems. The details of the Galileo system may change as its design evolves, but the general principles of its design are likely to remain unchanged.

All of these systems are designed to have a constellation of satellites that are deployed so that several satellites of the constellation are always visible from any point on the surface of the Earth. About 24 satellites, in circular orbits whose radii are about 26 000 km (4.2 Earth radii), are needed to implement this requirement. (The GPS constellation is the only one that actually realizes this capability at the present time. The GLONASS system does not have a full complement of active satellites at present, and the Galileo system is still in the design stage.)

## 2. The simple pseudo-range

I will discuss the signals transmitted by each satellite in detail below, but for now consider that each satellite continuously transmits a signal whose carrier frequency and modulation are derived from a single on-board frequency standard, which is usually a caesium or a rubidium device. (Although GPS satellites contain several atomic clocks, only one of them is active at any time, and the rest are kept in reserve as spares.) The modulation format includes a signal that is logically equivalent to a periodic ‘tick’. The transmitted signal also contains additional information including an estimate of the satellite orbit, the epoch of the transmission (determined by counting cycles of the on-board frequency standard) and an estimate of the relationship between the on-board clock and a system-wide average timescale (to be discussed below).

A receiver measures the physical time difference,  $\Delta T$ , between the tick received from the satellite and the corresponding tick from a local clock. In first order, this measurement is modelled as arising from several sources: the geometric path delay between the receiver at position  $(x_r, y_r, z_r)$  and the satellite at  $(x_s, y_s, z_s)$ , the delay through the receiver itself,  $\delta_r$ , and the offset between the satellite clock and the local clock,  $\delta t_{rs}$ :

$$\Delta T = \frac{\sqrt{(x_r - x_s)^2 + (y_r - y_s)^2 + (z_r - z_s)^2}}{c} + \delta_r + \delta t_{rs}. \quad (1)$$

Although equation (1) has units of time, it is called the pseudo-range, i.e. the geometrical range delay that would correspond to the measured time difference if the signal had travelled through a vacuum, the transmitter and receiver clocks were synchronized and the delay through the receiver were 0. (The significant corrections for the additional path delays due to the refractivities of the ionosphere and the troposphere are discussed below.)

The time difference in equation (1) above is measured with respect to the clock in the satellite. This raw measurement is useful in some limited situations. For example, it can be used for common-view measurements. The fact that parameters of the satellite clock are not known in this case does not matter, since the satellite clock cancels (at least to first order) in the common-view subtraction anyway. It can also be used to evaluate the satellite clocks themselves. However, most receivers add the offset between the satellite clock and the system time to this value and report the time difference between the local clock and the satellite system time,  $\Delta T_s$ :

$$\Delta T_s = \Delta T + \delta_s, \quad (2)$$

where  $\delta_s$  is computed from the message transmitted by the satellite. (The ground control stations compute satellite system time as a weighted average of the clocks in the satellites and the ground

stations, where the weights are determined from the past performance of each clock. The offset between this average system time and the clock in each satellite is estimated by the ground control stations, and parameters that characterize this offset are periodically uploaded into the satellite and are then retransmitted to users as part of the navigation message. The parameters broadcast by the satellite are predictions based on past performance, and their accuracy depends on the time since the last parameter update and the stability of the satellite hardware.)

The messages transmitted by the satellite contain another parameter that is often useful for time distribution. This is an estimate of the offset between GPS time and UTC(USNO), the real-time realization of coordinated universal time (UTC) maintained at the US Naval Observatory. This parameter is estimated based on measurements made at the Naval Observatory; the data are transmitted to the GPS control stations where they are uploaded into the satellite. This parameter can be used to provide traceability between the time broadcast by the GPS satellites and UTC, which is an international timescale maintained by the International Bureau of Weights and Measures (BIPM). To provide traceability to UTC, a user would combine these data with the offset of UTC(USNO) from UTC. This offset is published by the time section of BIPM in its monthly Circular T [3]<sup>2</sup>. These offsets are not needed for common-view applications, since the time of satellite clock cancels anyway. They are also not needed in purely geodetic applications; as we will see below these applications depend only on the relationship between the satellite clock and the system time and are independent of the relationship between GPS time and any other timescale.

The coordinates of the satellite at the time of transmission are determined from parameters in the broadcast message. If the location of the receiver is known *a priori* and if its delay has been measured in some ancillary experiment then equations (1) and (2) determine the time offset (in first order) between the local clock with respect to the satellite clock and with respect to the average system time, respectively. The additional calculations to estimate the offset with respect to UTC(USNO) or UTC can be incorporated as well, if necessary.

If the coordinates of the receiver are not known *a priori*, then it is possible to determine both the offset of the local clock and the full three-dimensional position of the receiver by tracking four satellites simultaneously. These four measurements can be used to solve for four unknowns: the three coordinates of the receiver, ( $x_r$ ,  $y_r$ ,  $z_r$ ), and the time offset,  $\Delta T_s$ , between the local clock and the system time. (An important special case is determining the coordinates of a point that is known to be on the surface of the Earth. This knowledge provides a relationship among the three coordinates of the receiver that can be used to reduce the number of satellites that must be tracked.) The accuracy of this solution will depend on the accuracy of the broadcast orbits and on the consistency of the set of  $\delta_s$  parameters transmitted by the different satellites being tracked. All of these parameters are extrapolations based on previous data obtained at the various tracking stations, so that the performance of the system depends on the stability of the satellite orbits and clocks and on the frequency with which the parameters broadcast by the satellite are updated.

The multi-satellite method, which is used when the position of the station is not known *a priori*, also depends on the fact that the delay through the receiver either is the same for all satellite signals or else is a known function of the satellite being tracked. The magnitude of any constant and satellite-independent delay is not important in determining the position of the station since it is absorbed into the clock solution, which is usually ignored in geodetic applications anyway. Constructing a geodetic solution based on data from different GLONASS satellites is potentially more difficult to do, since the satellites transmit at different frequencies and the delay through the receiver may be frequency dependent.

<sup>2</sup> See this reference for Circular T and for a description of the operation of the time section of the BIPM.

Although time transfer and geodetic position determination are often thought of as different applications, this is not always the case. For example, time transfer to a moving platform requires a simultaneous solution for both the position of the platform and the time offset of the clock. If the clock is moving rapidly or if it is not near the surface of the Earth, then additional corrections may be necessary to account for the Doppler shift due to the rapidly moving platform.

### 3. Coordinate systems and measurement frames

We must specify a coordinate system in order to evaluate equations (1) and (2). If we consider a receiver at a fixed location on the Earth, then the position of that user is normally specified in an Earth-centred, Earth-fixed (ECEF) system. This is a particularly convenient choice because the coordinates of a stationary receiver are constants in this system and it is straightforward (at least in principle) to relate these coordinates to the location of the receiver in other coordinate systems, such as the ones that are used in surveying and other ground-based civilian applications.

An example of such a system would be an ECEF system defined in terms of the WGS-84 model [4], which models the Earth as an ellipsoid whose major axis is in the equatorial plane of the Earth. Cross-sections of the Earth parallel to the equatorial plane are circular, while those normal to the plane are ellipsoidal. The  $z$ -axis of this model is normal to the equatorial plane and points towards the geographic North pole. In an ellipsoidal cross-section that contains the  $z$ -axis, the semi-major axis has a value of  $a = 6\,378\,137$  m, which is equal to the mean equatorial radius of the real Earth. The minor axis corresponds to the polar diameter of the Earth, and the magnitude of the semi-minor axis is  $b = 6\,356\,752.3142$  m. The model can also be specified in terms of an eccentricity,  $e$ , and a flattening,  $f$ , given by

$$e = \sqrt{1 - \frac{b^2}{a^2}} = 8.18 \times 10^{-2}$$

and

$$f = 1 - \frac{b}{a} = 3.35 \times 10^{-3}, \quad (3)$$

respectively. The  $+x$ - and  $+y$ -axes of the coordinate system are defined as pointing along the direction of  $0^\circ$  longitude and  $90^\circ$  East longitude, respectively, so that the coordinate axes form the usual right-handed system.

The WGS-84 Earth model is used by the GPS. The Russian GLONASS system uses a similar coordinate system called PZ-90 [5]. The model is basically the same as WGS-84, although the values of the parameters are slightly different. For example, the semi-major axis is 1 m smaller than the WGS-84 value, and the other parameters are also slightly different. There is no simple relationship between the coordinates of a given point in the two coordinate systems (or between the two system times, for that matter).

The International Earth Rotation Service (IERS) was established in 1987 to provide realizations of two coordinate systems: the International Celestial Reference System (ICRS) and the International Terrestrial Reference Frame (ITRF) [6]. The IERS uses data from many different techniques, including GPS observations, very long baseline radio interferometry (VLBI), and satellite laser ranging to define an ECEF coordinate system, and to measure the locations of reference stations with respect to it. The detailed operation of the IERS is described on its web page at [www.iers.org](http://www.iers.org).

The ‘geodetic’ position of a receiver is expressed in terms of its latitude and longitude with respect to the reference meridians of a model and its height above the reference ellipsoid.

This geodetic height is the minimum distance between the user and the reference ellipsoid, and this straight line distance does not point towards the centre of the Earth unless the user is at the Equator or at the poles. This geodetic height can be quite different from the height above mean sea level, which defines the geoid. The quantities that parameterize the satellite orbits are also transmitted in this system. Although an ECEF frame is convenient for users because the coordinates of a stationary receiver are constants with respect to its origin, it is rotating and is, therefore, not an inertial frame.

It is easiest to transform the coordinates of both the space vehicle and the receiver to an Earth-centred inertial (ECI) frame—one that is fixed in the centre of the Earth, but not rotating with respect to the ‘fixed’ stars, since light signals travel in straight lines with velocity  $c$  (in vacuum) in such a frame. (This choice ignores the small radial and tangential accelerations of the Earth in its orbit.) There are an infinite number of such frames, all of which differ from each other by rotations about the coordinate axes; all of these frames are conceptually equivalent, and the choice among them is based on computational convenience. At any instant of time, it is always possible to find an inertial frame in which the coordinates of the receiver are identical to their values in the standard ECEF frame, and this choice is often used since it usually simplifies the transformations that are required. This equivalence is only valid for an instant of time, since the ECEF is rotating whereas the equivalent ECI frame is not. Although this instantaneous equivalence is simple in concept, the details can become complicated in multi-channel receivers.

If a receiver is tracking multiple satellites simultaneously in order to solve for its position, then there are two ways of measuring the various pseudo-ranges: (1) either the receiver uses a tick transmitted at essentially the same instant from all of the satellites in view, in which case the different pseudo-ranges result in the measurements being made at different reception times or (2) the receiver makes all of the measurements at a given instant (as determined by its clock), in which case the signals were transmitted by the satellites at different epochs. The second method is conceptually easier, especially if the receiver is not stationary or if its local clock has a significant offset from the satellite timescale. A natural choice in this case would be to choose an ECI frame whose  $z$ -axis coincides with the polar axis of the WGS-84 system and which coincides with the ECEF frame of the receiver at the instant the measurements are made. The positions of the satellites could then be transformed into this frame using simple rotations about the chosen  $z$ -axis. Since the signals left the satellites at different times, a separate transformation may be needed for each satellite.

#### 4. Relativistic effects

In addition to the requirement that measurements associated with clock synchronization be performed in an inertial reference frame, there are other relativistic effects that affect satellite transmissions: the first- and second-order Doppler effects, the gravitational frequency shift and the effect of eccentricity in the orbit of the satellite.

The magnitudes of these effects can be calculated once the radius of the orbit of the satellite is given, and they will, therefore, vary somewhat from one design to another. The orbital radius of the GPS satellites is 26 600 km (about 4.2 Earth radii), and the orbital speed is, therefore,  $3.87 \text{ km s}^{-1}$ ; the GLONASS satellites have orbits of 25 460 km and their orbital speed is about  $3.95 \text{ km s}^{-1}$ . The Galileo satellites proposed by the Europeans will probably have a somewhat smaller orbital radius of about 24 000 km; the orbital speed will be correspondingly greater. I will use the GPS orbital parameters in the following calculations; the corresponding GLONASS and Galileo values would not differ significantly.

#### 4.1. The first-order Doppler shift

The magnitude and sign of this correction depend on the projections of the instantaneous velocities of the satellite and the observer onto the vector between them, where these velocities must be evaluated in the frame used for the analysis. This effect must be estimated by the tracking loop of the receiver in real time, since it changes the apparent frequencies of the received signals. To estimate the maximum magnitude of this effect, imagine a satellite travelling due West directly towards an observer on the Equator. The velocity of the receiver,  $v_r$ , is about  $450 \text{ m s}^{-1}$  towards the transmitting satellite—about 12% of the speed of the transmitter, denoted by  $v_t$ . Since both velocities are small compared to  $c$ , the first-order Doppler shift results in a fractional frequency shift (relative to the transmitted frequency) of approximately

$$\frac{\Delta f}{f} = \frac{v_r + v_t}{c - v_t} \approx 1.4 \times 10^{-5}, \quad (4)$$

where I have dropped terms proportional to quadratic and higher powers in  $v/c$ . The GPS carrier frequencies are about 1.5 GHz, so that the maximum first-order Doppler shift would be about  $\pm 19.4 \text{ kHz}$ .

The next term in the expansion would have a magnitude of order  $(v/c)^2$  times the value in equation (4), and this contribution might not be negligible in some situations. Using a fractional frequency shift of  $10^{-15}$  as a somewhat arbitrary threshold of significance, the quadratic term must be considered for velocities greater than about  $8.3 \text{ km s}^{-1}$ . As a practical matter, the contribution of the uncertainty in the velocity to the linear term in equation (4) is usually much larger than the size of the quadratic correction.

#### 4.2. The second-order Doppler effect

This effect depends on the orbital speed of the satellites, and is therefore the same for all satellites with the same orbital radius. It produces a fractional frequency shift whose magnitude is given by  $\gamma$ , where

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}} \approx 1 + 8.3 \times 10^{-11}, \quad (5)$$

and where  $\beta = v/c$ . The effect is always a ‘red’ shift—the clock in the satellite appears to have a lower frequency compared to a reference clock at the location of the observer (whom we imagine to be on the rotating geoid).

#### 4.3. The gravitational frequency shift

This effect depends on the difference in the gravitational potential at the source and the observer. If we take a clock on the rotating geoid as the reference, then the clock in a satellite appears to this observer to be running faster by about

$$\frac{\Delta f}{f} = \frac{\Delta\phi}{c^2} = -\frac{GM}{c^2} \left( \frac{1}{4.2R_e} - \frac{1}{R_e} \right) \approx 5.3 \times 10^{-10}, \quad (6)$$

where  $M$  and  $R_e$  are the mass and radius of the Earth, respectively,  $G$  is the gravitational constant and  $c$  is the speed of light. This effect is also the same for all satellites in a given orbit.

The second-order Doppler shift and the gravitational frequency shift are of opposite sign. Taken together, they would result in an observer on the geoid measuring a net fractional frequency increase of  $4.47 \times 10^{-10}$  relative to an observer in the satellite. The proper frequency of the oscillator on each satellite is adjusted downward by this amount so that

the signals received on the geoid appear to have their nominally correct coordinate value. For the GPS, where the nominal frequency of the reference oscillator is 10.23 MHz, the actual proper frequency of the oscillator is lower by about 4.573 MHz. Although this seems like a small correction, if it were not applied the received time would diverge from its nominal value by more than  $38 \mu\text{s day}^{-1}$ —an enormous offset rate in the time and frequency business.

Note that this correction is applied so that an observer on the rotating geoid sees the clocks running at the nominal frequency. This is the standard reference frame in the time and frequency business, but it is not an inertial frame, nor is it the obvious best choice if (or when) space travel becomes common.

#### *4.4. The eccentricity of the satellite orbit*

The eccentricity contributes in two ways to the frequency as observed at the receiver. In the first place, the eccentricity results in a varying radial acceleration which is equivalent to a varying gravitational potential. In the second place, the eccentricity results in a varying orbital speed which affects the second-order Doppler shift. These effects vary together with the period of the satellite orbit. The apparent frequency of the satellite clock is lower at perigee and higher at apogee. This frequency modulation is removed by the code and carrier tracking loops in the receiver, and the magnitude of the effect is estimated based on the broadcast ephemeris.

### **5. The index of refraction of the transmission path**

The signals broadcast by satellite systems do not travel in a vacuum, and their velocities are less than  $c$  as a result. This effect is usually small enough, and the elevation angle of the path with respect to the horizon is usually large enough that the signals from the satellites can be considered as travelling along geometrical straight lines with a speed given by  $c/N$ , where  $N$  is the index of refraction of the medium. Since the index of refraction is very nearly 1, it is more common to speak of the refractivity,  $n$ , where  $n = N - 1$ .

The assumption that the paths are geometrical straight lines actually implies a limit on the spatial gradient of the index of refraction rather than on its absolute magnitude. The spatial gradient can be quite large near the surface of the Earth, so that using simple geometrical straight lines is often not adequate in modelling ground-to-ground transmissions. In many cases, the atmosphere can be approximated as composed of a series of horizontal layers of different refractivities, so that the spatial deviation of the signals due to changes in the refractivity becomes smaller as the elevation angle increases even when there are large variations in the refractivity at the boundaries between the layers. It is quite common to restrict observations to satellites with elevation angles of  $20^\circ$  or more for this reason. The bending effect is often not negligible even for signals transmitted by satellites at these elevation angles, and this is considered below in the discussion of obliquity factors.

The path from a satellite to a ground station can usually be divided into two parts: the upper part through the ionosphere, where the refractivity is dominated by the contribution of the free electrons, and the lower part where the refractivity is dominated by the contribution of dry air, with smaller contributions from water vapour and other trace constituents. The refractivities of the two portions of the path are discussed in the following sections.

#### *5.1. The effect of the ionosphere on the path delay*

The refractivity of the path from a satellite to a receiver through the ionosphere is a function of position and of the frequency used to transmit the information. The dependence on frequency

is *dispersion*, and the frequency is identified with a subscript on the refractivity. The variation of the refractivity with position is written as  $n(r)$ , where  $r$  is a parameter that measures the line-of-sight distance between the transmitter and the receiver. As we will show in the following discussion, it is possible to estimate the refractivity of the ionosphere by measuring its dispersion, i.e. by measuring the difference in the apparent pseudo-ranges at two different carrier frequencies. The two different frequencies that are used for this purpose are called  $L_1$  and  $L_2$ . The GPS and GLONASS systems use different pairs of frequencies for this purpose, but the concept of estimating the refractivity from the measured dispersion is the same for both systems.

### 5.2. The multiple wavelength method

If the refractivity can be expressed as a product of two functions, one of which depends only on the frequency of the signal (which we call  $F$ ) and the second of which depends only on the properties of the transmission medium (which we call  $G$ ), then it is possible to determine the refractivity of the transmission medium by measuring the time dispersion—the apparent time difference between measurements made using two different frequencies. This idea can be realized in two ways: the dispersion measurement can be used to correct the data at one frequency, or the two measurements can be used to construct a ‘refractivity-free’ measurement.

If the length of the physical path between the transmitter and the receiver is  $D$ , then the transit times measured using two different carrier frequencies  $f_1$  and  $f_2$  will be

$$t_1 = \int \frac{(1+n_1(r))}{c} dr = \frac{D}{c} + \frac{F(f_1)}{c} \int G(\dots) dr \quad (7)$$

and

$$t_2 = \int \frac{(1+n_2(r))}{c} dr = \frac{D}{c} + \frac{F(f_2)}{c} \int G(\dots) dr, \quad (8)$$

where the quantities  $n_1(r)$  and  $n_2(r)$  are the refractivities at the two frequencies at the coordinate point  $r$  along the path. (The refractivity is the difference between the actual index of refraction at a point and 1, which is its vacuum value.) The second form on the right-hand side of these equations makes use of the fact that the refractivity at each frequency can be decomposed into a known function  $F$  that is independent of the path and depends only on the frequency of the carrier and a second known function  $G$  that depends only on the parameters of the medium such as its density, pressure, temperature, etc. The integrals must be evaluated along the transmission path. This would be difficult or impossible for a real-world path through the atmosphere, since the required parameters are not known and cannot be easily measured. However, since the integral of  $G$  is independent of the carrier frequency, we can eliminate  $G$  in terms of the difference in the two measured transit times.

$$\int G(\dots) dr = c \frac{t_1 - t_2}{F(f_1) - F(f_2)}. \quad (9)$$

Substituting equation (9) into equation (7),

$$t_1 = \frac{D}{c} + (t_1 - t_2) \frac{F(f_1)}{F(f_1) - F(f_2)} = \frac{D}{c} + (t_1 - t_2) \frac{n_1}{\Delta n}. \quad (10)$$

The first term on the right-hand side is the time delay due to the transit time of the signal along the path of physical length  $D$ , and the second term is the additional delay due to the refractivity of the medium. (The second term can never be negative, since it is the product of two terms both of which always have the same sign.)

The second term depends only on the measured time difference between the signals at the two frequencies and on the known frequency dispersion of the refractivity. It does not depend on the physical length  $D$  itself or on the details of the spatial dependence of  $G$ , and

we would have obtained exactly the same result if only a small piece of the path  $D' < D$  was actually dispersive, while the rest of the path had a non-zero refractivity but no dispersion. (Note the implicit requirement that the signals at both frequencies travel along the same physical path,  $D$ . This requirement is usually satisfied except for optical signals transmitted over ground to ground paths longer than about 50 km. The atmospheric dispersion can steer the two signals in this configuration by enough so that the two beams diverge by more than the characteristic size of a cell of constant refractivity. The two beams then sample different refractive indices, and the assumptions of the calculation are no longer valid.)

The fact that the difference in the arrival times between signals of different frequency can be used to estimate the refractivity of the path is an important conclusion, because it means that a receiver can estimate the delay through the ionosphere by measuring the difference in the arrival times of identical messages sent using two different carrier frequencies. The  $L_1$  and  $L_2$  frequencies transmitted by both the GPS and GLONASS satellites are used for this purpose.

The additional time delay due to the refractivity of the ionosphere is given in first order by [7]

$$\Delta t = \frac{40.3\rho_e}{cf^2}, \quad (11)$$

where  $c$  is the velocity of light,  $f$  is the signal frequency in Hz and  $\rho_e$  is the total electron content integrated over the path through the ionosphere in units of electrons m<sup>-2</sup>. The magnitude of the total electron content varies in time and in space. Typical values range from about  $10^{16}$  to  $10^{19}$ . Using  $\rho_e = 10^{16}$  and  $f = 1$  GHz, the additional time delay due to the refractivity of the ionosphere is about 1.3 ns. The total electron content can vary by a factor of 50 or even 100 during the solar day, with a maximum that occurs usually in the early afternoon (local time). Especially near sunrise and sunset, this diurnal variation can introduce a significant azimuthal dependence to the effect of the ionospheric refractivity.

Using equation (10) with the two GPS signal frequencies  $L_1 = 1575.42$  MHz and  $L_2 = 1227.60$  MHz, we find

$$\frac{n_1}{\Delta n} = \frac{L_2^2}{L_2^2 - L_1^2} = -1.54, \quad (12)$$

where the negative sign results from the fact that the refractivity decreases with increasing frequency and is therefore smaller at  $L_1$  than at  $L_2$ . Thus we also expect that  $t_1 < t_2$ .

Instead of substituting equation (12) into equation (10) and using the measured time difference resulting from the dispersion to correct the measurement at  $L_1$ , we can substitute equation (12) into equation (10) and solve for the physical time delay  $D/c$ :

$$\frac{D}{c} = 2.54t_1 - 1.54t_2, \quad (13)$$

where the quantity on the right-hand side of equation (13) is the ‘ionospheric-free’ measurement, and the corresponding pseudo-carrier frequency is often referred to as  $L_3$  in the literature.

Although equation (13) provides a measurement of the time delay that is independent of the ionosphere (to first order, limited by the validity of equation (11)), these data will be noisier than a single-frequency measurement using only  $L_1$ . Assuming that the signal to noise ratio is the same for both frequencies (which is often not true), the measurement noise at  $L_3$  should be poorer than at  $L_1$  by a factor of about  $\sqrt{(2.54^2 + 1.54^2)} \approx 3$ . This trade-off between signal to noise and accuracy is particularly important for common-view measurements (in which two stations observe the same satellite at the same time). The ionospheric correction is often ignored for short baselines in common-view applications, since the effect will cancel in

common-view anyway and there is no need to pay the price of the degradation in the signal to noise ratio that results from using the ionosphere-free combination. The correction is more important over longer baselines, since the effects of the ionosphere tend to be different at the two stations—especially near local sunrise and sunset at either site.

The refractivity and the dispersion of the ionosphere become negligible at optical frequencies. The correction is usually ignored even at microwave frequencies above 10 GHz. Although this would suggest that life would be easier at these higher microwave frequencies, the effect of the troposphere becomes larger because of the presence of water vapour lines, especially above 20 GHz. Conversely, the dispersion between the two carrier frequencies would be larger (and therefore easier to measure) as the frequencies are moved further apart. This would complicate the design of the receiver, and might also make it more difficult to protect the frequency allocation from interference by other ground-based uses.

The height of the ionosphere is about 200–500 km, but there is appreciable diurnal and seasonal variation in this value. Since the orbital radius of a satellite is usually much larger than these values, the full effect of the ionosphere must be considered in a satellite to ground configuration, while the effect of the ionosphere usually can be ignored in satellite to satellite transmissions. This might not be true in all circumstances. Transmissions to and from a low-flying satellite might be affected by only part of the ionosphere, and the effect might vary during the orbit. The electron density in the various layers of the ionosphere varies by several orders of magnitude both with altitude and with the time of day. The diurnal variation is usually largest in the ‘E’ layer, which has a height of about 100 km. Typical electron densities might be  $10^9$  electrons  $\text{m}^{-3}$  during the night and  $5 \times 10^{11}$  electrons  $\text{m}^{-3}$  during the day. The variation with altitude is equally large: typical daytime values would be about  $10^8$  electrons  $\text{m}^{-3}$  at an altitude of 50 km, increasing by a factor of 10 000 at 200 km. This spatial and temporal variation would have to be folded into the detailed orbital parameters in order to estimate the effect in a particular configuration. These variations do not invalidate the concept of estimating the refractivity from the dispersion, but they introduce a temporal and spatial variation into the coefficients of equation (11) that complicates practical analyses.

This variation becomes more complicated if the receiver is moving. Both frequencies will be Doppler shifted in this case, and these shifts must be removed by the tracking loop in the receiver. If the dispersion is measured as a phase difference between the two carriers, then the relationship between phase and time will be modified as a result of these Doppler shifts. In addition, if the motion is through the dispersive portion of the path, the velocity modifies the size of the dispersion (through the effective value for the total electron content in the case of the ionosphere). This effect is most likely to be significant for satellite to satellite transmissions when the two satellites are nearly on opposite sides of the Earth. It may be difficult in practice to separate changes in the measurement dispersion due to temporal changes in the ionosphere from changes caused by uncertainties in the position of one of the satellites or due to differential cycle slips between the  $L_1$  and  $L_2$  phase meters in the receiver.

The total electron content of the ionosphere at any point is usually specified as a value at the zenith. This value must be corrected for an obliquity factor if the satellite is not directly overhead. An estimate of this factor for a receiver on the Earth is presented below; the model we will present is unlikely to be adequate for satellite to satellite transmissions, and a direct integration of the density of free electrons along the actual path is likely to be necessary.

### 5.3. Group and phase velocities

The previous discussion considered the effect on the ionosphere of two frequencies,  $L_1$  and  $L_2$ . In fact, most receivers do not use these carrier frequencies themselves, but rather the codes

which are modulated on these carriers. The propagation speed of these codes is characterized by the group index of refraction, which is not the same as the phase index of refraction that characterizes the speed of the carrier. Using the first-order functional form for the refractivity given above in equation (11), the group and phase refractivities have the same magnitude but opposite sign. The group index of refraction uses the magnitude in equation (11) with a positive sign, so that the effective group velocity is less than  $c$ , while the phase index of refraction uses the magnitude with a negative sign, so that the effective phase velocity is greater than  $c$ . The calculation above, which is based only on the magnitude of the refractivity, is valid for either index. In order to distinguish measurements made using the code with those based on the carrier, the former are usually identified by  $P$  rather than  $L$ . Thus  $P_1$  is often used to designate the pseudo-range measurement made using the code transmitted using the  $L_1$  carrier frequency, etc.

Since the group and phase delays will have opposite dependences on the total electron content of the ionosphere, it is possible to estimate the time variation of the total electron content of the ionosphere using a single frequency receiver that measures the variation in the differential delay between the code and the carrier of a GPS signal. Although this method has been tested by a number of groups, it is difficult to implement in practice because of the ambiguities in calibrating the effective delay through a GPS carrier-phase receiver. This effective delay is modified by cycle slips; for many common receiver designs it also changes each time the receiver is re-started.

#### 5.4. The troposphere

The refractivity of the troposphere at radio frequencies is given by [8]

$$n = 77.6 \times 10^{-6} \frac{P}{T} + 0.373 \frac{e}{T^2}, \quad (14)$$

where  $P$  is the total atmospheric pressure in millibars,  $T$  is the temperature in kelvin and  $e$  is the partial pressure of water vapour in millibars. The refractivity is independent of frequency, and the group and phase indices of refraction are therefore the same. The lack of dispersion means that the refractivity cannot be measured using two frequencies as in the case of the ionosphere discussed above. For typical values ( $P = 1000$  mb,  $T = 300$  K,  $e = 10$  mb),  $n = 300 \times 10^{-6}$ . Using these parameters, the optical length exceeds the physical length by about  $30 \text{ cm km}^{-1}$  of path length, which corresponds to an increase in the travel time of  $1 \text{ ns km}^{-1}$ .

The first term in equation (14) contributes about 85% of the total value. The ratio  $P/T$  is proportional to the density of dry air, and the integrated effect of the first term on a nearly vertical path, therefore, can be modelled pretty well (usually to within a few per cent) using values of  $P$  and  $T$  at the surface (assuming that the atmosphere is in thermodynamic equilibrium, which makes the calculation possible but is hardly ever exactly true). The effect of the second term is smaller, but it is much harder to measure, since it usually varies in an irregular way with time, azimuth and elevation. (Also note that it does not depend on the density alone and so cannot be estimated using a single point measurement, even if we assume that the distribution is in thermodynamic equilibrium.)

Some work has been done on measuring the water vapour contribution using radiometers. A number of different types of radiometers have been tested, but the most common type estimates the integrated density of water vapour (along a single line of sight) by comparing the received power at one of the water vapour lines with the power at a nearby frequency [9, 10]. The frequencies that are used for this comparison are usually on the order of 25–30 GHz. The sensitivity of the method depends on the noise temperature of the receiver and its surroundings, while the accuracy depends on the calibration of the power measurements. This method is not widely used because the instruments are expensive and can be difficult to calibrate. A second

type of radiometer exploits the dispersion between an optical and a microwave frequency [11]. This method is potentially more accurate, but it is also much more complicated because the dry air also contributes to this dispersion, and the two effects must be separated using ancillary measurements.

Because of these difficulties in estimating the effect of tropospheric water vapour, the second term is often ignored at locations that have low humidity. If the atmosphere is not isotropic or is not in thermal equilibrium, its contribution is often smaller than the uncertainty in the dry-air part of the refractivity.

The integrated effect of the atmosphere is to increase the path by about 2 m (7 ns) in the zenith direction. If the atmosphere is isotropic in azimuth and elevation, then this excess path delay will vary as the cosecant of the elevation angle, reaching a value of about 25 m (83 ns) at an elevation of 5°. Many analyses model the atmosphere in this way, using a single value for the zenith delay and scaling this parameter by the cosecant of the elevation angle. (A more complete discussion of this effect follows below.) This zenith delay can be obtained using a real measurement, but is more usually estimated (at least in post-processed analyses) from the data themselves by fitting a cosecant function to a long-arc track. As with the ionosphere, these effects are usually not important in satellite to satellite transmissions, but this might not be true for a satellite in a very low orbit or when two satellites are on opposite sides of the Earth.

### 5.5. Characterizing the troposphere using optical measurements

The dispersion of the troposphere is significant at optical frequencies, and it would be possible to measure the refractivity in this way. Each component of the atmosphere makes a different contribution to the overall refractivity. The simplest models ignore trace constituents (the most important of which is usually CO<sub>2</sub>), and use only two components to model the refractivity: one due to dry air and the second due to water vapour. Thus,

$$n = F_a(f)G_a(P, T) + F_w(f)G_w(e, T), \quad (15)$$

where the subscript ‘a’ refers to the dry-air contribution, the subscript ‘w’ refers to the contribution due to water vapour, and the other symbols are as defined above. The functions  $F$  and  $G$  are quite complicated. (Note that atmospheric scattering and scintillation make it impractical to use the phase of the optical carrier directly, and measurements are always made using some form of modulation. The indices of refraction presented below are, therefore, the group values.) Instead of using the frequency of the signal in these expressions, it is more common to use the wave number, which is the reciprocal of the wavelength in micrometres and is therefore directly proportional to the frequency. The relationship between the wave number in inverse micrometres and the frequency  $f$  in Hz is approximately

$$\sigma = \frac{f}{3 \times 10^{14}}. \quad (16)$$

(The approximation in equation (16) comes from rounding the value that is used for the speed of light to  $3 \times 10^8$  m s<sup>-1</sup>). In these units, optical frequencies have wave number values of about 1.5 or 2. Using these parameters, approximate expressions for the parameters in equation (15) are [12, 13]:

$$F_a(\sigma) = 23.71 \times 10^{-6} + 6389.4 \times 10^{-6} \frac{130 + \sigma^2}{(130 - \sigma^2)^2} + 45.5 \times 10^{-6} \frac{38.9 + \sigma^2}{(38.9 - \sigma^2)^2}, \quad (17)$$

$$F_w(\sigma) = 64.87 \times 10^{-6} + 1.74 \times 10^{-6} \sigma^2 + 0.03 \times 10^{-6} \sigma^4 + \dots, \quad (18)$$

$$G_a(P, T) = \frac{P - e}{T} \left[ 1 + 57.9 \times 10^{-8} - \frac{9.3 \times 10^{-4}}{T} + \dots \right], \quad (19)$$

$$G_w(e, T) = \frac{e}{T} + \frac{e^2}{T} - 8.7 \times 10^{-7} \frac{e^3}{T} + 8.3 \times 10^{-4} \frac{e^4}{T^2} + \dots \quad (20)$$

Using the same values for  $P$ ,  $T$  and  $e$  as above, the refractivity of the atmosphere is  $285 \times 10^{-6}$  in the red (633 nm) and  $298 \times 10^{-6}$  in the blue (442 nm). Using these values for the optical refractive indices, the factor that multiplies the time dispersion in equation (10) is

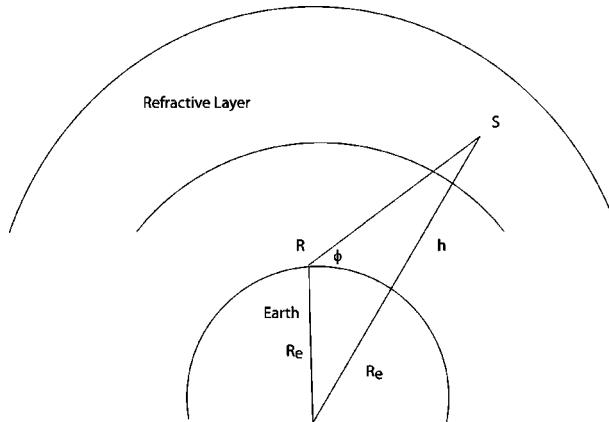
$$\left| \frac{n_{\text{red}}}{n_{\text{red}} - n_{\text{blue}}} \right| \approx 22. \quad (21)$$

This factor is much larger than the corresponding value for the microwave dispersion of the ionosphere given in equation (12) above, and it means that the uncertainty in a two wavelength measurement of the atmospheric refractivity is dominated by the uncertainty in the measurement of the dispersion. This measurement is made even more difficult by the fact that light at blue wavelengths is more strongly attenuated by atmospheric scattering and absorption than a lower-frequency red signal would be, so that practical dispersion measurements are usually limited by the characteristics of the blue signal generator. It is also possible to construct an ‘atmospheric-free’ analogue of equation (13). Since the value of the expression in equation (21) is much larger than unity, the uncertainty of the resulting estimate is very nearly a factor of 22 times less accurate than the uncertainty of a single-wavelength measurement. This is an expensive penalty, and multiple wavelength optical systems are not commonly used for this reason.

### 5.6. Obliquity factors

The corrections for both the ionosphere and the troposphere are normally specified in terms of the increment in the path delay due to the refractivity of the medium when the satellite is at the zenith. These values must be corrected using an obliquity factor for observations made when the satellite is at a lower elevation angle. The simplest approach is to assume that the refractive medium is both homogeneous and isotropic. This assumption tends to be poorest for the troposphere, especially at stations where the nearby topography has significant relief.

If we assume that the ionosphere and troposphere are homogeneous concentric shells of refractive material (see figure 1), then the obliquity factor for a satellite whose elevation angle



**Figure 1.** Obliquity factor assuming that the ionosphere is a spherical shell at a height  $h$  above the Earth. The receiver, located at position  $R$  is receiving a signal from the satellite at position  $S$ . The parameter  $R_e$  is the radius of the Earth.

is  $\phi$  (with respect to an observer on the ground) is given by

$$\Delta = \frac{R_e + h}{\sqrt{(R_e + h)^2 - R_e^2 \cos^2 \phi}}, \quad (22)$$

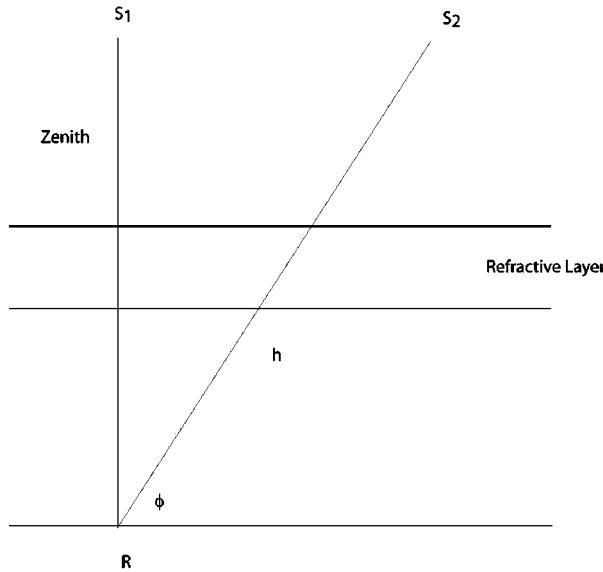
where  $R_e$  is the radius of the Earth and  $h$  is the average altitude of the refractive layer, measured relative to a station on the surface of the Earth. This expression multiplies the zenith value for the refractivity. The fraction gives the ratio between the zenith path length and the path length along the straight line path to the satellite.

Since the thickness of the troposphere is much smaller than the radius of the Earth, the troposphere is often modelled using  $h = 0$ , as shown in figure 2. In this case the expression reduces to

$$\Delta = \frac{1}{\sin \phi}. \quad (23)$$

This simpler expression could also be obtained using a model in which the Earth is flat and both the ionosphere and troposphere are horizontal layers with a very small refractivity. The obliquity factors computed using these two expressions differ by about 5% at an elevation angle of 45°, and this difference increases as the elevation angle decreases. Neither expression works very well at elevation angles less than about 20° (because spatial inhomogeneities are usually much important at these low elevation angles), and many receivers can be programmed to ignore satellites whose elevation is this low. The problem at low elevation angles is especially serious in models of the troposphere, since there is usually significant turbulence and spatial inhomogeneity near the surface of the Earth.

Neither of these obliquity factors is adequate if the refractivity is not azimuthally symmetric (i.e. if the refractivity is not symmetric about the zenith direction). It is possible to remedy this deficiency by means of more complex functions which include an azimuthal dependence. One such function would multiply the factors given in equations (22) and (23) by an additional term proportional to the cosine of the azimuthal angle. This additional multiplicative term



**Figure 2.** Simplified model for the obliquity factor, assuming that the Earth is flat and the refractive layer is a horizontal slab.

would be of the form

$$1 + a \cos(\theta - \theta_0), \quad (24)$$

where  $a$  and  $\theta$  are the amplitude of the azimuthal variation and the azimuthal angle, respectively. The amplitude,  $a$ , and orientation of this function with respect to the ECEF coordinate axes,  $\theta_0$ , must be deduced from the data.

## 6. Detailed parameters of each system

In the previous sections, I presented the considerations that describe time transfer using any of the satellite-based systems that I am considering. In the next sections I will describe the detailed parameters of the GPS and GLONASS constellations. I will also discuss the design for the Galileo system, although the details may change as the system design evolves.

### 6.1. GPS Orbits and signal formats

The GPS constellation consists of 24 satellites in six orbital planes, which are inclined at an angle of 55° with respect to the Equator. The radius of the orbit is 26 600 km, the orbital speed is 3.87 km s<sup>-1</sup> and the orbital period is about 12 h. As seen by an observer on the Earth, the satellites return to very nearly the same point in the sky once every sidereal day (23 h, 56 min).

The primary oscillator on each satellite operates at a frequency of 10.23 MHz. (As I mentioned above, this is the frequency as seen by an observer on the Earth. The proper frequency is lower by about 5 MHz.) This frequency is multiplied by 154 to produce the  $L_1$  carrier frequency at 1575.42 MHz and by 120 to produce the  $L_2$  carrier frequency at 1227.60 MHz. Both carriers are modulated by a pseudo-random sequence with a chipping frequency of 10.23 MHz (the ‘P’ code), and the  $L_1$  carrier is also modulated by a slower pseudo-random sequence with a chipping frequency of 1.023 MHz (the C/A code). (The signal used to modulate the  $L_2$  carrier can be changed by the control stations. The three possibilities are: P code with data, P code with no data and C/A code with data; the data message is described in the next paragraph.)

In addition, a navigation data message is transmitted on  $L_1$  and  $L_2$  at 50 bit s<sup>-1</sup>. This navigation message is transmitted on  $L_1$  and  $L_2$  by combining it with both C/A code and P code data streams using an ‘exclusive or’ operation. This means that a ‘1’ bit in the data stream reverses the sign of the faster codes while a ‘0’ bit in the data stream leaves them unchanged. The resulting bit streams modulate the carriers using a bi-phase shift key method. The P code + data stream is modulated in phase quadrature with the C/A code + data stream on the  $L_1$  carrier. The  $L_2$  carrier normally has only one modulation signal and the second phase is not modulated. Using the frequencies defined above, there are 20 460 C/A chips and 204 600 P chips between each data bit. Increasing the data rate would make the code acquisition more difficult since the potential code reversals would be more frequent. However, the information that is transmitted in the data messages could be increased by the same factor.

In addition, when ‘anti-spoofing’ mode (AS) is activated (which is the normal state of affairs at present), the P code is encrypted, and the encrypted version is called the Y code. The encryption makes the code unusable by non-authorized receivers, but does not change the chipping rate. In addition to no change in the chipping rate, many other aspects of the GPS signal also are not changed by this encryption. The clear and encrypted versions of the code are often referred to by the single identifier P(Y) code for this reason.

The access to all of the signals can also be degraded for non-authorized users by encrypted dithering of the frequency of the primary oscillator. This is called selective availability (SA). In addition, SA can also be used to introduce an encrypted offset error into the navigation message.

This introduces an error into the pseudo-range as calculated by a non-authorized user. Selective availability is currently turned off, and its future is not known. The frequency dither introduced when SA is on degrades using the signals for determining either position or time, but it can be cancelled almost completely by using common view, since the dither is common to the signals received at all of the stations and, therefore, cancels in the pairwise difference. When the receiver can track more than one satellite at the same time, the cancellation is most effective when the subtraction is performed separately for each satellite that is being tracked, since the SA dithers applied to different satellites are uncorrelated. In general, this means that common-view subtraction is more effective for time comparisons than for position determination, unless the stations are quite close together so that they are all tracking the same group of satellites. In general, errors in the navigation message due to SA would not be cancelled in common view, since the dither usually affects the pseudo-ranges to the various stations by different amounts; this type of dithering was not observed when SA was on in the past.

The  $L_1$  and  $L_2$  transmissions are used to estimate the delay through the ionosphere using the two-wavelength dispersion technique described in the previous section. Since the  $L_2$  frequency normally is modulated only by the encrypted Y code (plus data), a non-authorized user cannot directly measure the  $L_1 - L_2$  dispersion to estimate the delay added by the ionosphere. However, it is possible to measure this dispersion by noting that the P code + data bit stream is the same on both  $L_1$  and  $L_2$  and by looking for a peak in the cross-correlation between these two signals. This cross-correlation method can succeed even when the code is encrypted. One way of doing this is to compute the cross-correlation using the squares of the code streams. This effectively removes the navigation message without the need to actually parse it.

*6.1.1. The GPS pseudo-random codes.* Since all of the GPS satellites transmit using the same pair of frequencies, identifying them must be based on the transmitted code. This arrangement is called code division multiple access (CDMA). The concept is simple: the codes transmitted by each satellite are orthogonal to each other, so that the receiver can identify any satellite by constructing a replica of its code and computing the cross-correlation between the received signal and the local copy. The process is a bit more complicated in practice, because each satellite also has a unique Doppler frequency offset, which must also be measured. Therefore, a real receiver must have two tracking loops that are linked: one that locks on the Doppler-offset carrier and a second that locks on the code. The cross-correlation process that locks on the code is further complicated by the bits of the navigation message. A ‘1’ bit in this message reverses the sign both of the pseudo-random code and the cross-correlation with its local replica, while a ‘0’ bit leaves the sign unchanged. (Since  $(-1)^2 = (+1)^2 = 1$ , this reversal can be removed by squaring the signals.)

Each of the pseudo-random noise (PRN) codes is a ‘Gold’ code [14]. The C/A codes are synthesized by combining the outputs of two 10 bit shift registers using an exclusive-or operation. The shift registers are clocked at the chipping frequency of the code. The first register, G1, has taps to feedback stages 3 and 10 to the input. The output of the second generator, G2, is delayed relative to the first one by a number of chips, unique to each satellite and derived from the satellite ‘PRN number’. The code generators in all of the satellites are the same except for this delay, so that specifying this number is sufficient to identify the code and the satellite that transmitted it. (In addition to the PRN number, which characterizes the transmitted codes, satellites also have a space vehicle (SV) number, which is assigned by the controllers when it is launched. The two numbers are not necessarily identical.)

The P codes are synthesized using four 12 bit shift registers. These registers are combined in pairs using exclusive-or logic, and the resulting pair of bit streams are handled in a manner analogous to the C/A generator described above.

The C/A code is 1023 bit long and repeats every 1 ms. The P code is about 38 weeks long, but each satellite is assigned a unique week-long subsection of it. The P code transmitted by each satellite repeats every week starting from midnight Saturday night/Sunday morning (as measured by the clock on the satellite).

**6.1.2. Using the GPS timescale.** Since the P code is normally encrypted, most users have access only to the C/A code and the navigation message, and the following description assumes that the user does not have access to the P code.

The cross-correlation of the received C/A code with a local replica generated by the receiver produces a series of ticks at 1 ms intervals which are synchronous with the transmitted code, which is driven by the oscillator on the satellite. The navigation message is extracted by using the fact that it can reverse the sign of the cross-correlation every 20th tick. After the receiver has locked on the code and the carrier, it begins to assemble the bits of the navigation message by looking for these phase reversals. These phase reversals specify the ticks of the 50 Hz navigation message clock.

The bits in the navigation message are assembled into 30 bit words. Since the message is transmitted at  $50 \text{ bits s}^{-1}$ , each word requires 600 ms to transmit. These words are assembled into groups of 10. Each group of 300 bit is called a sub-frame and requires 6 s to transmit.

The first two words in every sub-frame are the telemetry word (TLM) and the handover word (HOW), and the receiver looks for these words to synchronize the phase of the navigation message clock with respect to the corresponding clock in the satellite. The TLM is identified by its 8 bit preamble: 1000 1011, and the remainder of the word contains parameters used by authorized users. The first 17 bit of the HOW contain the most-significant 17 bit of the epoch which will occur at the start of the next following sub-frame. The full time parameter is 19 bit long and is measured in units of 1.5 s relative to the start of the GPS week (midnight of Saturday/Sunday). The value transmitted in the HOW is, therefore, the time relative to the start of the week in units of 6 s (four 1.5 s intervals). The TLM/HOW combination has significant redundancy to assist the receiver in identifying it. The preamble is always the same, and consecutive times in the HOW differ by one (except during the rollover on Saturday night as described below). Furthermore, both the TLM and HOW words contain parity-check bits, which are calculated by the satellite and can be checked by the receiver to verify that the word has not been corrupted in transmission.

A time value of 1 is transmitted in the HOW at the start of the week (midnight Saturday/Sunday). It identifies the time of the next sub-frame as 6 s after midnight. (The full 19 bit time parameter at that point will have a value of 4.) The maximum value is 100 799, which corresponds to a time of 604 794 s (6 days, 23 h, 59 min, 54 s), and points to the start of the last 6 s frame of the GPS week. This maximum value is transmitted 12 s before midnight on the following Saturday night. The next HOW, which is transmitted at 6 s before midnight, contains the value 0, which signals that another GPS week is about to begin (at the start of the next sub-frame).

In addition to the 19 bit time parameter, which measures the GPS epoch relative to the start of each week, there is an additional 10 bit week count, which measures the number of weeks since the origin of the GPS timescale. It is incremented by 1 each time the GPS time rolls over to zero. The original origin of this counter was midnight 5 January–6 January 1980. However, the maximum value of this parameter is only 1023, and it rolls over to 0 each time this maximum value is reached. The first rollover occurred at midnight 21 August–22 August 1999, and a new origin was established at that time. Additional rollovers will occur about every 19.7 yr, so that the GPS timescale is fundamentally ambiguous at this level.

The impact of this ambiguity on a geodetic receiver is generally limited to the time period during and immediately following the rollover. The first potential difficulty is that all of the satellites will not transmit the rollover at the same instant, because the rollover is determined by the clock in each satellite, and these clocks may differ by up to 1 ms. Even if the satellite clocks were perfectly synchronized, the difference in the times of flight between different satellites and a receiving station could make a contribution as large as 10 ms to the difference in the times the signals were received. The second potential difficulty is that the position of the satellite and the offset of its clock from GPS time are transmitted as functions of time, and the receiver will have to take the rollover into account in evaluating these functions when the extrapolation period crosses the rollover. (A smaller version of this problem occurs every Saturday night when the time value in the HOW word rolls over to 0, and some receivers have had problems at this time as well.)

Since the relationship between GPS time and the time kept by the rest of the world does not enter into a position solution, the effects described above tend to be transients that disappear once the rollover is over and the transmitted ephemeris parameters use a post-rollover reference time. In the extreme situation, a permanent error of 19.7 yr in parsing the GPS time would have no impact at all on position solutions.

The impact of the rollover on a clock display would also be limited. Although the epoch would be wrong by 19.7 yr, both the day of the week and the time of day would be correct, and users who did not compute a civil date from GPS time would not see the problem.

The rollover problem has been addressed in a number of different ways. One solution uses a hard-coded date in the receiver software (the date it was compiled, for example) to estimate the number of rollovers that have occurred since 1980. Another solution tries to estimate the year by looking at the number of leap seconds transmitted in the GPS message and noting that (at least at present) there has been somewhat less than one leap second per year. Neither of these solutions is really adequate for an application that uses GPS time to assign an epoch to a datum for archival purposes, since these hints on estimating the rollover value are generally not preserved.

*6.1.3. The GPS navigation message.* In addition to providing synchronization signals for the receiver clock through the HOW and TLM, the navigation message contains other parameters that are important for time and frequency users. The parameters are identified by the sub-frame ID code, which is transmitted in bits 20–22 of the HOW. The following paragraphs describe the parameters that are most useful for time and frequency users.

Sub-frame 1 contains the GPS week number, 6 bit describing the health of the satellite, an estimate of the  $L_1 - L_2$  dispersion due to the ionosphere (which can be used by single-frequency receivers), and the coefficients of a polynomial that estimates the difference between the satellite clock and GPS time. The polynomial coefficients give the predicted time offset, frequency offset and frequency ageing with respect to an origin time that is also part of the message. Any error in this prediction enters directly into the error budget for a position solution using several satellites, since multiple satellite pseudo-ranges can be related only through GPS time. The error is less serious for a common-view observation (see below), since that technique cancels or attenuates offsets in the satellite clock. The receiver must deal with the possibility that a week rollover has occurred since the transmitted origin time. The implication of the documentation is that these parameters will be updated at least once during each GPS week because there is no provision for detecting a double week rollover.

Sub-frames 2 and 3 contain the ephemeris parameters, which are used by the receiver to compute the position of the satellite. As with the time parameters in the first sub-frame, this

computation is with respect to a transmitted origin time, and the receiver must deal with the possibility that a rollover has occurred since that epoch.

The ephemeris for each satellite is described using six Keplerian elements. These parameters would be strict constants of the orbit in the simple ‘two-body’ case, which assumes that the gravitational potential of the Earth can be represented as a simple  $1/r$  function (where  $r$  is the radius vector from the centre of mass to the satellite) and that other perturbing influences (gravitational effects of other bodies, solar radiation pressure, etc) are not present. The actual orbit is still described using the same elements, but they are no longer strict constants of the motion. Instead they vary slowly with time, and this variation must be estimated by the tracking stations and inserted into the navigation message that is broadcast by each satellite.

The first three Keplerian elements define the shape of the orbit. They are: the length of the semi-major axis of the ellipse,  $a$ , the eccentricity,  $e$ , and the time,  $\tau$ , at which the satellite crosses the perigee, which is the point on the orbit nearest the Earth. (The semi-major axis crosses the ellipse at perigee.) The GPS actually uses the mean anomaly instead of the time at perigee; the definition of this parameter in terms of the ‘true’ anomaly is discussed below.

The ‘true’ anomaly at any instant is the angle (conventionally measured counter-clockwise) between the direction of perigee defined above and the actual position of the satellite at that instant. This true anomaly varies linearly with time if the orbit is exactly circular. However, it has a more complex variation for an elliptical orbit. The changing radius vector in an elliptical orbit combined with conservation of angular momentum implies a variable angular velocity. The mean anomaly is defined so that it varies linearly with time with a variation that agrees with the variation of the true anomaly averaged over a complete orbit. The relationship between the mean anomaly,  $M$ , at some time  $t$  and its value at some reference time  $t_0$  is given by

$$M - M_0 = \sqrt{\frac{\mu}{a^3}}(t - t_0), \quad (25)$$

where  $\mu$  is the reduced mass of the Earth/Satellite system multiplied by the gravitational constant,  $G$ . By definition, both the mean and the true anomalies are 0 at the time of passage through the point of perigee, so that the relationship between the mean anomaly at the reference time  $t_0$  and the time of perigee passage is given by

$$M_0 = -\sqrt{\frac{\mu}{a^3}}(\tau - t_0). \quad (26)$$

The quantity  $\sqrt{(\mu/a^3)}$  plays the role of an angular velocity in these expressions, and is called the mean motion of the satellite, usually represented by  $n$ . From this definition, the orbital period of the satellite is simply  $2\pi/n$ .

The second three Keplerian elements define the orientation of the orbit in space with respect to the ECEF coordinate system in which the  $x$ - $y$  plane is the equatorial plane of the Earth and the  $z$ -axis is perpendicular to this plane. The three parameters are then the inclination of the orbit,  $i$ , which is the angle between the normal to the plane of the orbit and the  $z$ -axis,  $\Omega$ , the longitude (or right ascension) of the ascending node (the point at which the orbit crosses the equatorial plane moving upward towards positive  $z$ - and the  $x$ -axes) measured in the equatorial plane, and  $\omega$ , the angle between the ascending node and the direction of perigee measured in the orbital plane.

These parameters are not quite constants of the motion because of the various additional perturbations mentioned above. The navigation message contains estimates of the time derivatives of these parameters and other information to permit the receiver to estimate the actual position of the satellite between message updates.

Sub-frames 4 and 5 contain an almanac giving a set of reduced-precision ephemeris parameters for the other satellites in the constellation. Once the signal from a single satellite has been parsed, these data enable a receiver to compute which other satellites to look for.

Finally, sub-frame 4 includes parameters that relate GPS time to UTC(USNO) and also gives information regarding current and future scheduled leap seconds. The leap seconds are added to UTC but not to GPS time, so that the difference between the two timescales changes by 1 s each time a leap second occurs. As we mentioned above, the leap second parameter can be used to estimate the approximate value for the civil year, which is useful for detecting a rollover back to zero of the GPS week.

The difference between GPS time and UTC(USNO) is a slowly varying function of time whose magnitude is usually less than 10 ns. This parameter is needed only by users who need traceability to UTC(USNO) using a one-way signal. It is not needed for a position solution or for users who use the common-view method for time distribution (see below).

The accuracy of GPS time (i.e. the absolute magnitude of the difference between it and UTC(USNO)) is not important for many users. For example, the accuracy of a position determination depends on the accuracy of the offset transmitted by each satellite between its internal timescale and GPS time and not on the accuracy of the system time itself with respect to UTC(USNO). In the same way, the accuracy of the traceability to UTC(USNO) depends on the accuracy of the transmitted value for the offset between GPS time and UTC(USNO), and there is no absolute requirement for the magnitude of this parameter to be small. While it is not an absolute requirement for the operation of the satellite system, the fact that it is kept small (by steering GPS time to UTC(USNO)) is a convenience for many users, who can approximate UTC(USNO) using GPS time. On the other hand, the stability of the GPS time is important because this stability defines the accuracy of the extrapolation for a given interval between message updates. Alternatively, improving the stability can be used to decrease the number of messages updates needed to realize a given level of performance. Therefore, the method of steering GPS time to UTC(USNO) must be something of a compromise between time accuracy, which would favour more aggressive steering, and frequency smoothness, which can be compromised if the steering adjustments are too large.

## 6.2. GLONASS orbits and signal formats

The GLONASS system is similar to GPS in general concept. The full constellation consists of 24 satellites arranged as eight satellite slots in each of three orbital planes which are spaced 120° apart along the Equator. The slots are defined based on their mean anomaly (the angular position of the slot with respect to the Earth's Equator measured along the orbit). The orbits are circular with a radius of 25 460 km and are inclined 64.8° with respect to the Equator. The period is about 11 h, 15 min, so that each satellite appears to move into the adjacent slot every day, and returns to its initial apparent position after every 8 sidereal days (approximately 191 h, 28 min). Eight satellites are operating as of May 2002, but one of them has been declared temporarily unusable.

The GLONASS P code has a chipping frequency of 5.11 MHz and a repetition period of 1 s; the C/A code chipping frequency is 511 kHz and repeats every millisecond. Both codes are pseudo-random sequences generated using methods analogous to those used in the GPS. The GLONASS C/A code has one-half the chipping rate and one-half the length of its GPS equivalent. It is thus easier to find by exhaustive search, but has only one-half the resolution. Similarly, the shorter length and more rapid repetition period for the GLONASS P code relative to its GPS counterpart mean that while it is easier to acquire the GLONASS P code in principle, the code will have somewhat poorer correlation properties. Since the GLONASS P code is

much shorter than its GPS counterpart, a HOW, which is used by a GPS receiver to acquire the P code after it has locked onto the C/A code and the bit clock of the data message is not really necessary. The GLONASS P code can be acquired by exhaustive search, since there are only of order 10 offsets to test once the receiver has locked onto the C/A code.

Unlike the GPS, all satellites transmit the same pair of codes, but the carrier frequencies are different (frequency division multiple access (FDMA)). These frequencies are given by

$$f_1 = 1602 + 0.5625k \text{ MHz} \quad (27)$$

and

$$f_2 = 1246 + 0.4375k \text{ MHz}, \quad (28)$$

where  $k$  is an integer that is assigned to each satellite. In the initial design,  $k$  was an integer from 0 to 12 that was unique to each satellite. The resulting frequencies interfered with radio astronomy observations, and a lower set of frequencies, including negative values of  $k$  (down to  $-7$ ) and eventually up only to  $+4$ , will be phased-in over the next few years. Neither the original range of integers nor the new lower-frequency set contains enough values to assign a unique one to each satellite, so some sharing of frequencies will be needed if the full constellation of satellites is deployed. One possibility would be to assign the same value for  $k$  to satellites on opposite sides of the Earth. This would not present a problem for ground-based observers, since they could not see both of them simultaneously, but satellite-based receivers would require additional processing (similar to the tracking loops in GPS receivers) to distinguish between the two signals.

These rather widely spaced frequency allocations complicate the design of the front-end of the receiver, especially the temperature stability of the filters that are usually needed to reject strong out-of-band interference. The calibration of GLONASS receivers is both more difficult and less accurate because of this; in some early multi-channel receivers the effective delay varied from channel to channel even for the same carrier frequency. On the other hand, cross-talk between signals from the different satellites is much less of a problem; interference from a single-frequency source is also less of a problem with GLONASS, since the interfering signal is less likely to affect the signals from the entire constellation.

Apart from the requirement for a tunable front-end, the operation of a GLONASS receiver is basically the same as for GPS hardware, and there are commercial units that can track satellites from either constellation. The only practical difficulty with tracking satellites from both constellations at the same time is that the GLONASS system uses the PZ-90 reference frame instead of the WGS-84 frame used by GPS, and the GLONASS system time is traceable to international standards through UTC(SU), rather than through UTC(USNO).

## 7. The Galileo system

A consortium of European agencies have begun working on a satellite constellation similar to GPS that will be called Galileo. Although the system is still in the design phase, the preliminary configuration will have 30 satellites in three orbital planes. The radius of the orbit will be about 24 000 km. This is somewhat less than either GPS or GLONASS, so that the orbital period will be correspondingly shorter. Unlike both GPS and GLONASS, which were initially designed for military applications, the initial system design of Galileo emphasizes civilian applications, including interfaces to emergency locator beacons, public transport, etc. The constellation may also include one or more geostationary satellites for real-time monitoring of the active constellation. The system time will be referenced directly to International Atomic Time (TAI), but the details of how this is to be done are not yet finalised.

## 8. Civil navigation overlay

The function of this overlay is to complement the existing satellite systems by providing additional services. Three most important of these services are:

- (a) evaluating the health of each satellite in the GPS and GLONASS constellations and transmitting this health information in real time to help users avoid using unhealthy satellites for navigation;
- (b) transmission of additional GPS-like ranging signals to provide improved reliability and redundancy;
- (c) transmitting differential corrections to GPS and GLONASS signals to permit users to cancel or mitigate the effects of errors in the broadcast orbits and the satellite clocks.

The focus of these efforts is to support real-time precision navigation, including automatic landing of aircraft and similar applications. The combination of these overlay services is called Wide Area Augmentation System (WAAS). Unlike GPS and GLONASS satellites, which have transmitters and code generators in the satellite, the geostationary satellites in this system will carry transponders that rebroadcast signals that originate at a ground control station.

The constellation will consist of four (eventually five) geostationary satellites. The signals broadcast from these satellites will be similar to the GPS format and will be transmitted at a single frequency (GPS  $L_1$  at 1575.42 MHz). The transmissions will be synchronized to GPS time so as to emulate a GPS satellite. In addition to the usual C/A code, message symbols are transmitted to provide the additional information outlined above. The symbol rate of 500 symbols per second is synchronized with the 1 kHz repetition rate of the GPS C/A code. The C/A code broadcast by these satellites will use PRN numbers starting with number 120; the code generators will use G2 delays that produce codes that are orthogonal to the codes generated by ‘real’ GPS satellites (see the discussion of GPS codes above).

The navigation overlay is much more important for assisting applications that depend on position determination than it is for applications involving time and frequency distribution. The reason is that once the position of a receiving station is known, a multi-channel receiver has the ability to make several independent estimates of the time difference between the local clock and GPS time by using different satellites. Using these independent measurements to detect an outlier is often realized in a receiver autonomous integrity monitoring (RAIM) algorithm. This procedure usually can detect outliers much more effectively than can be done using WAAS transmissions from a remote station. Both the local clock and much of the receiver hardware are common to the measurements using different satellites in a single receiver, and a bad satellite stands out more clearly. (A RAIM algorithm can also be used to detect and remove a bad satellite in a position solution, but it usually takes at least five satellites in view to be able to do this reliably.)

## 9. Measurement strategies

### 9.1. *The single station method*

Although there are a number of variations, there are two basic strategies for using a satellite system such as GPS to distribute time and frequency information: single-station methods and common-view methods. (I am assuming in this discussion that the receiving stations are fixed with respect to an ECEF frame and that the coordinates of the stations in that frame have been determined by a previous measurement. If the station is moving or if its coordinates are not known *a priori*, then the simple time-transfer methods described here would not be appropriate, since both the offset of the station clock and its coordinates would need to be determined as

part of the satellite solution.) The choice between these two strategies is a trade-off between performance and complexity, although many of the most important errors and problems are present in both techniques. I will discuss these methods in detail using the GPS as an example, but the general techniques would be equally applicable to GLONASS or Galileo. (In principle, these strategies might also be implemented using a satellite constellation that included a mix of the different types of satellites. This is not practical at present because the GPS and GLONASS coordinate systems and time references are different, and solving for these differences in the receiver is not a trivial job.)

The single-station method is conceptually very simple: a user at a known location measures the difference between the station clock and the time code as received from one or more satellites. Since the coordinates of the station are known, the user can compute the pseudo-range to each satellite being tracked once each broadcast ephemeris message has been parsed. For the GPS, this requires reading sub-frames 1, 2 and 3. Once the hardware has measured the time difference between the local clock and the transmitted time code, the data in the broadcast message can be used to compute the difference between the station clock and GPS time or UTC(USNO). If the receiver can track several satellites at once, then the data from each one can provide an independent estimate of either of these differences, and it is possible to detect a bad satellite or a measurement outlier by comparing these differences using a RAIM algorithm. The sensitivity of a RAIM algorithm will depend on any satellite-dependent (or receiver-channel-dependent) offsets, especially if those offsets are time varying. The largest effect of this type is usually multi-path reflections of the primary signals by objects near the receiver, although frequency-dependent variations in the channel delays can also be a problem in the GLONASS system. An impedance mismatch at either end of the cable between the antenna and the receiver will result in reflections in the antenna cable which have many of the same characteristics as atmospheric multi-path reflections.

The receiver can estimate the effect of the ionosphere using either the measured  $L_1 - L_2$  dispersion (if the receiver can process both wavelengths) or using the model of the ionosphere that is part of the broadcast message. As we discussed above, the two-wavelength dispersion data generally will be noisier than the single-wavelength data by about a factor of 3. Most users do not correct for the refractivity of the troposphere, primarily because there is no easy way to estimate the size of this effect.

The single-station method is sensitive in first order to any error that affects the pseudo-range, including errors in the broadcast ephemeris, in the model of the ionosphere and in the refractivity of the troposphere. If the user is trying to recover GPS time or UTC(USNO), then any errors in the broadcast estimates of those parameters also enter directly into the error budget. During the 1990s, however, the largest contribution to the error budget by far for a non-authorized user was the intentional clock dither imposed as part of SA. The amplitude of this dither was of order 100 ns—in practice, this was about a factor of 10–15 larger than the combined contributions of all of the other problems mentioned above. The frequency dither of the satellite clocks due to SA is currently turned off, and its future use is unclear. At least for now, the fact that it is turned off means that one-way non-authorized users can realize stabilities that were previously obtainable only with common-view methods.

## 9.2. *The common-view method*

In the common-view method, two stations agree to observe the same satellite at the same time and to process the measurements in the same way. (Many stations follow the Technical Directives published by the BIPM to acquire and analyse common-view data. The details of these directives are discussed in more detail below.) The basic measurement strategy is the

same as in the single-station case, with the additional step of subtracting the data from the stations point by point after each observing session is finished. Assuming that the two stations have comparable noise levels, this subtraction implies that such measurements will always be noisier than the underlying single station data by a factor of  $\sqrt{2}$ . The advantages are worth the price in most cases, since the data are rarely limited by white phase noise anyway.

The primary advantage of the common-view method is that many of the effects that limit the accuracy of the GPS transmissions are cancelled or significantly attenuated by the subtraction at the end of the observations. This cancellation is nearly perfect for stations that are reasonably close to each other, since the effects of the troposphere and the ionosphere tend to affect both paths in the same way. The same thing is true for the parameters in the broadcast ephemeris—especially the clock model parameters and the parameters that define the radial distance between the stations and the satellite. Single frequency (i.e.  $L_1$  only) receivers are usually adequate for common-view measurements over relatively short baselines (up to several hundred km long), since the effect of the ionosphere tends to be the similar on both paths so that it cancels in the difference. There is no point in paying the increase in the noise of about a factor of 3 that results from processing the two wavelengths at each station. There is often no advantage in incorporating the broadcast model of the ionosphere for the same reason.

A second advantage of the common-view method is that it can provide traceability<sup>3</sup> to a particular National Metrology Institute (NMI) or timing laboratory, when that is an important requirement. A number of NMIs and timing laboratories (including NIST) offer this type of service, which provides traceability to the official timescale of the laboratory using common-view comparisons between the user and the timing laboratory. These comparisons are substantially independent of the performance of the GPS itself.

The common-view method was particularly powerful in dealing with the effects of SA, since subtracting the observations at the two stations cancelled the satellite clock almost perfectly, thereby removing the effects of the frequency dither of the satellite clock.

Although the common-view method is useful in cancelling or attenuating many of the sources of noise in the GPS, it is important to keep in mind that it does not provide any improvement for problems that are not common to the receiving stations. There are a number of significant sources of error that are in this category, including the delay through the receiving equipment (and especially its variation with temperature) and the effect of multi-path reflections caused by a time-varying combination of the direct signal from the satellite with other signals that are first reflected from objects near the antenna.

## 10. Measurement details

Since many frequency standards produce pulses at 1 Hz, most timing receivers are designed to make a measurement of the time difference between the local clock and the received time code at this rate. The 1 Hz pulses from the receiver are derived from the clock that drives the C/A code correlator. The fundamental C/A code correlation process produces a series of pulses at the repetition rate of this code, which is 1 kHz. The pulse from this series that is associated with the transition to the next second is chosen based on the HOW in the data stream as described previously.

If the receiver can track several satellites simultaneously, then there will be several such pulses, which differ because of the noise in the measurement processes in each channel and

<sup>3</sup> By ‘traceability’ I mean that there is an unbroken chain of measurements between the signal received by an end-user and a national or international standard. Each measurement in the chain has an associated uncertainty. The adequacy of any chain of measurements is normally evaluated based on the requirements of the end-user, and some methods may be deemed adequately traceable only by some users (or only for some applications).

the systematic offsets associated with each satellite. Multi-channel receivers generally use the data from all of the satellites in view to construct a composite 1 Hz tick; the individual satellite data that are used to construct this composite are transmitted on a separate output channel so that the times of the ticks from each satellite that is being tracked can be computed. Not all multi-channel receivers produce these data—in some cases only the composite tick and its epoch are output.

Since the common-view method depends on each station observing the same satellite at the same time, a multi-channel timing receiver that produces only a composite output tick cannot be used for applications requiring strict common-view processing. There are two exceptions to this limitation. The first is when two receivers are close enough to each other so that they track the same ensemble of satellites. The composite output ticks from both receivers are derived from the same ensembles, and tend to have fluctuations that are well correlated as a result. The second application is where the primary need for common view is traceability rather than the highest possible accuracy. Although the two stations may be tracking different ensembles of satellites, the variation among the ensemble of satellites is small enough that it can be ignored. It is not possible to provide absolute values for the ranges of validity of these exceptions, but they will generally be satisfied for receivers separated by no more than a few tens of km or for applications requiring traceability with an uncertainty of about  $1 \mu\text{s}$  or greater.

The epoch associated with the tick produced by the receiver is computed from the data messages received from each satellite. Timing receivers generally use the UTC second as the time tag, so that the received epoch must be converted to UTC using the parameters described in the previous sections. It is possible for there to be some ambiguity in this conversion in a multi-channel receiver, since the parameters in the messages from different satellites may not be completely consistent with each other. These discrepancies can arise, for example, because the data messages in the various satellites that are being tracked were uploaded at different times. One strategy would be to use the most recently uploaded data message that is available, but not all receivers do this. Some receivers use the simpler strategy of always using the data message from the lowest-numbered channel that is actively tracking a satellite.

Since the input from the local clock is usually just a 1 Hz tick, the measured time differences are expressed, at least initially, modulo 1 s. There is an inevitable discontinuity in the output data, which can be either at 0 and 1 s or at  $\pm 0.5$  s depending on the design of the receiver. In principle, this discontinuity can be removed for subsequent measurements by keeping track of the previous data so as to detect a rollover through the discontinuity, but this function is generally performed by the analysis programs rather than by the receiver itself.

These 1 Hz measurements have significant white phase noise, and averaging these data to reduce the contribution of this noise is almost always the optimum strategy. (This was true even in the days when SA dominated the variance at somewhat longer averaging times.) This strategy obviously remains optimum as long as the data continue to be characterized by white phase noise. Typical averaging times in practice are about 15 or 30 s. The BIPM technical directives specify that fifteen 1 s measurements are to be averaged by fitting them with a quadratic function of time using conventional least squares, implying that even the variance at this level has significant coherence. The value of the fitting function at the midpoint of the interval is then used in the next step of the averaging process, as described below. Although this is the standard method used by all timing laboratories, it is not necessarily better than other methods that are simpler to implement and require fewer computing resources to evaluate. In experiments conducted at NIST, a least-squares fit using a simpler linear function and a straight average with a simple outlier detector worked about as well as the procedure outlined in the BIPM Technical Directives.

Whatever method is chosen, deterministic variations in the pseudo-range become significant for averaging times longer than a few seconds, so that the details of the processing are going to affect the answer. It is particularly important to standardize these details, especially if the full common-view cancellation of the satellite clock and errors in the broadcast ephemeris is to be realized.

To further complicate the problem, early receivers intended for common-view service could track only one satellite at a time, and they needed some time to receive the full ephemeris message after switching to a new satellite. Since the full ephemeris message requires 12.5 min to be received, the standard observation period was initially set to 13 min to guarantee that every receiver would have had a chance to read the full ephemeris message during the track time. This 13 min track length is still used by all NMIs, timing laboratories and their customers, even though it is less important for multi-channel receivers and may be longer than optimum in some cases.

## 11. The BIPM Technical Directives for averaging GPS data

All data that are to be used for international time and frequency coordination are averaged as specified by the BIPM (Bureau International des Poids et Mesures), and many timing laboratories also use the same averaging process in common-view calibrations with their customers [15]. We have already identified a number of aspects of this averaging process: the 13 min total track time and the quadratic least-squares fit to groups of fifteen 1 s measurements. There will be 52 groups of these 1 s measurements in the 13 min track time. The quadratic function that is used to fit each group is evaluated at the midpoint time of the corresponding group, and the resulting 52 values are fit again with a linear function of time using standard least squares. The final result is the value of this linear fit at the midpoint of the 13 min track time. Additional parameters (the slope of the line, the RMS of the residuals, etc) are also reported. The report format is also specified as part of the standard.

Tests at NIST have shown that while there is nothing very wrong with the analysis method presented above, it is not obviously optimum either [16]. The performance at very short times (on the order of a few seconds) tends to be dominated by white phase noise, and any method that averages this noise is adequate to the job. The performance at times longer than a few seconds tends to have deterministic variations, which are not necessarily well modelled by the quadratic and linear fits that are part of the standard, so that these fits do not necessarily produce unbiased estimates of the least-squares parameters. (This is still true today when SA is off, and it was particularly true in the days of SA, because the SA clock dither was a pseudo-random function that was not well modelled by any polynomial function of the time.) For example, an important problem is the systematic variation in the observations that results from multi-path reflections or from a small error in the assumed ‘known’ position of the receiving antenna. These variations interact with the standard fitting procedure in a complicated way that depends on the position and velocity of the satellite and the amplitude and phase (relative to the direct signal) of any reflections.

The common-view analysis process tends to work pretty well in spite of these limitations because the biases and inadequacies of the fitting procedure tends to produce systematic effects that are similar at all stations. In other words, the fact that all of the stations are doing the same thing is more important than the statistical robustness of the process itself. However, there would be a number of advantages to using shorter tracks with a simpler analysis method. Two of the most significant advantages would be the detection of outliers and the estimation of multi-path (which we discuss below). The increase in the size of the resulting data sets is

not nearly so serious today as it would have been 15+ yr ago when the current analysis method and track formats were developed.

The final aspect of the track schedule is the daily advance of 4 min in each of the track times. As we discussed above, the GPS satellites return to very nearly the same apparent position in the sky every sidereal day (23 h, 56 min), and a track schedule that is synchronous with the sidereal day automatically preserves the same common-view geometry from day to day. As we show in the next section, this method has the unfortunate side effect of making it more difficult to evaluate the magnitudes of multi-path reflections; this evaluation would be facilitated if shorter and more frequent tracks were used. Such a strategy would not have been feasible in the days of single-channel receivers, but it would be very straightforward today.

In the initial implementation of the receiver hardware, the tracking schedules had to be entered manually into each receiver, and a receiver that did not ‘know’ the tracking schedule could not participate in the common-view procedure. The reason for this is that combining a track length of 13 min with a daily advance of the start time of 4 min results in start times that have no simple relationship to each other or to the time the satellite actually becomes visible at a particular site. This was not a serious limitation in the days of single-channel receivers, since the receiver had to be told which of the several satellites that were visible at any epoch should be tracked. However, it is an unnecessary restriction on an all-in-view receiver.

The newer tracking schedules use a fixed 16 min grid of start times. Since the grid of tracking times advances 4 min every day, the schedule repeats exactly after 4 days, and a receiver that is synchronized to the origin time of this 16 min grid does not need to know the actual tracking schedule for its location. While not all 16 min slots are assigned to the tracking schedule of a particular location, every assigned track will have a start time that matches one of the 16 min slots. The advance in the track times of 4 min every day means that there are only four possible distinct 16 min grids, and the entire schedule can be calculated from a single start time on the reference day in the past. This newer system is particularly advantageous for multi-channel receivers, since it means that they do not have to be re-programmed each time a new tracking schedule is published. There is no particular advantage for single-channel receivers, since they have to be told which satellite to track in any case.

The combination of a predictable track schedule and the increased availability of multi-channel receivers has made possible a number of hybrid measurement strategies, which are somewhere between the single-station and common-view techniques described above. In one strategy, all multi-channel receivers are programmed to track all satellites in view (up to the capacity of the receiver) using 13 min tracks aligned to the 16 min grid specified above. Since all receivers are on the same time grid, it is a simple matter to find the satellites that were in common view between any pair of stations using a simple post-processing algorithm.

An alternative method is to combine the data from all satellites at each station, much as is done in a multi-channel receiver that produces only a single composite output pulse. These ‘melting-pot’ techniques estimate the time difference between the local clock and GPS time. A bad satellite or a measurement outlier can be detected using a RAIM algorithm as described above. These composite estimates can then be combined with the corresponding data from a second station. The result is not as good as same-satellite common view, since the two composite signals may be formed from different groups of satellites, but it can be considerably better than a simple one-way measurement, especially over baselines that are short enough that many of the same satellites are observed at both stations. This method was not effective against SA, which can only be cancelled by strict same-satellite common view.

## 12. Measurement uncertainties and offsets

Although errors in computing the pseudo-range and noise in the receiver contribute to the overall error budget of satellite time and frequency distribution, many systems are limited in practice by two effects that we have not discussed in detail: multi-path reflections and changes in the receiver calibration.

There are two types of multi-path reflections that are important.

1. Reflections of the received signal from objects that are near the antenna. These signals travel longer paths than the direct line-of-sight distance and therefore arrive at the antenna after the direct signal. The phase of the composite received signal is delayed relative to its line-of-sight value as a result. The magnitudes and phases of these reflections relative to the direct signal are complicated functions of the position of the satellite in the sky, the size and location of nearby reflecting objects, the sensitivity of the antenna to signals arriving from below or from the side and similar factors. Since the resolution of a typical GPS receiver is a few ns, which is on the order of 0.1% of a C/A code chip period, even weak reflections can produce significant offsets. Although the effects of multi-path reflections vary as the satellite moves across the sky and the geometry changes, it is difficult to see these changes using a receiver that follows the published BIPM tracking schedules. Since these schedules are intentionally synchronized to the sidereal day, the time-varying effect of multi-path reflections tends to be converted to a static time offset (generally different for each satellite) that varies only very slowly over several months as the approximate 4 min daily advance in the track time becomes less accurate in producing the same geometry day after day. The result is a slow change in the time-difference data which can be hard to distinguish from flicker and random-walk processes in the clock, changes in the effective delay through the receiver and other long-period errors.

2. Reflections in the cables that connect the receiver to the antenna and to the local reference clock. These reflections are produced by impedance mismatches at connectors or cable junctions. These problems are most serious in receiver designs where the cable between the antenna and the receiver carries information at the  $L_1$  and  $L_2$  carrier frequencies, and they are less serious if the signal is mixed with a local oscillator at the antenna and converted to a lower frequency for transmission to the receiver. Although this effect can be minimized by careful design, the complex impedance of the system components can be temperature dependent, so that it is effectively impossible to eliminate these reflections in a real-world environment where the temperature of the cable cannot be stabilized.

There are no easy solutions to these multi-path problems. Antennas with choke rings and ground planes and better antenna cables can help. Another solution would be to use a more directional antenna which had a smaller side-lobe response than the usual omni-directional design. This type of antenna could be realized by combining the signals from several antennas that are mounted on a common frame and are separated by not more than a few wavelengths at the  $L_1$  frequency. Different combinations could be used to track satellites at different azimuths simultaneously. The usefulness of this strategy would depend on how well the delay and gain of the different elements could be balanced. Multi-element antennas are still being developed, and multi-path effects often remain the most serious problem at many sites even when choke ring antennas and better cables are used.

The impact of reflections in the antenna cable can be minimized by installing an attenuator in the cable. The goal is to have any reflected signal pass through the attenuator 3 (or more) times, while the direct signal passes through only once. If the attenuator is installed very close to the receiver, then any reflection back into the receiver from the attenuator itself has only a small phase shift with respect to the primary signal. If the antenna cable is not too

long (30 m or less), many GPS receivers can detect signals even when a 30 db attenuator is inserted in the antenna cable. Some receivers require a booster amplifier to function in this configuration. Both this optional booster amplifier and the antenna itself may require power to function; this power is usually transmitted through the antenna cable itself, and the attenuator must be designed to pass this voltage from the receiver back up the cable to the active devices.

The calibration of the effective delay through satellite receivers is a second important factor that is difficult to address. Two calibration methods are used.

1. *Calibration using a satellite simulator.* A simulator is a device that generates signals that mimic real satellites. Since these signals are driven by a local reference clock in the simulator, the response of the receiver to these signals can be compared to the output of the reference clock to establish the effective delay through the receiver. A potential difficulty is that the test configuration may not include the real antenna and cable, so that the test conditions may not reflect the actual operating environment of the system. The antenna and its cable can be tested using a chamber specially designed for this purpose, but it is difficult to reproduce the exact outside environment in this way. A second difficulty with this method is that simulators are complex and expensive, and not many laboratories have them.

2. *Calibration using short-baseline common view.* In this method, the receiver to be calibrated is operated near a second standard receiver. Two antennas are placed near each other, and the two receivers are driven by a common reference clock. This method obviously tests the receiver in a more realistic environment, but it produces a relative (rather than an absolute) calibration, and even a small distance between two antennas can produce a significant difference in the responses of the two systems to multi-path reflections.

The calibration of GLONASS receivers is generally more difficult because of the FDMA format that is used by these satellites (see above). The front end of GLONASS receivers usually contain filters and other frequency-sensitive elements, and the delays through these components often vary with frequency and temperature. Different satellites may have different effective delays for this reason, and the whole business may be temperature sensitive in addition. A GLONASS receiver may therefore require a matrix of effective delay parameters, which specify the delay as a function of which satellite is being observed and which hardware channel is being used for the observation.

These effects of multi-path reflections and receiver calibration (and other effects that are a function of the hardware at a particular station) are not removed or attenuated in common view, and these effects tend to dominate the error budgets of common-view observations. A number of laboratories (including NIST) operate a number of different types of GPS receivers in parallel from the same reference clock. This strategy is obviously not perfect, but it provides a rough estimate of the magnitudes of these local effects.

### 13. Carrier-phase methods

The  $L_1$  and  $L_2$  carriers from every satellite are referenced to the same oscillator that produces the time codes. Since the frequency of the  $L_1$  carrier is about 1500 times the chipping frequency of the C/A code, a phase measurement of the carrier has potentially much greater resolution than the corresponding measurement using the time code. Carrier-phase methods have been used for many years in geodetic studies, and a number of groups are experimenting with using them for time and frequency distribution. The initial results are quite promising, although a number of problems remain. In the following sections, we describe time and frequency transfer using the phase of the carrier from GPS satellites; signals from other satellites could be used as well if the required infrastructure (precise orbits, etc) becomes available.

### 13.1. Carrier-phase model

The basic idea of the carrier-phase method is simple: instead of measuring the time difference between the 1 Hz ticks of the local clock and the corresponding events derived from the PRN code transmitted by the satellite, the receiver makes a phase measurement between an oscillator at the receiving station and the GPS carrier frequency [17]<sup>4</sup>. The phase difference between the received signal and the local oscillator at some epoch  $t$  is  $\delta\phi(t)$ . It is modelled in first order by

$$\delta\phi(t) = f^s \frac{r}{c} + f^s \delta^s - f_r \delta_r - (f^s - f_r)t + N, \quad (29)$$

where  $f^s$  is the transmitted carrier frequency (either  $L_1$  or  $L_2$ ),  $r$  is the geometrical distance between the position of the satellite (at the instant of transmission) and the receiver,  $\delta^s$  is the time offset between the satellite clock and the GPS time,  $f_r$  and  $\delta_r$  are the frequency offset and time offset of the oscillator in the receiver, respectively and  $c$  is the velocity of light. The integer  $N$  is the number of complete cycles in the phase difference and is included to make the computed phase difference have a value between 0 and 1.

The first term on the right-hand side of equation (29) models the advance in the measured phase due to the transit time of the signal from the satellite to the receiver, assuming that the path is in vacuum. This term must be evaluated in an inertial reference frame, and the previous discussion about how to choose this frame is applicable here as well. The second term models the advance in the measured phase due to the fact that the clock on the satellite has a time offset with respect to GPS time, and the third term is the corresponding phase offset resulting from the time offset of the receiver clock. The second and third terms are usually ignored, since carrier-phase analyses are always done in common view, and the offset frequencies that contribute to both terms are small. The fourth term results from the frequency difference between the two oscillators. Since the phase is only measured modulo a full cycle, the last term includes the unknown integer number of cycles, which must be determined during the post-processing. This unknown integer includes the integer parts of all of the other terms; both its initial value and any subsequent phase rollovers must be estimated during the post-processing analysis. This first-order model must have additional terms to model the refractivities of the ionosphere and the troposphere, which are computed using the same basic methods discussed above for a code-based measurement.

### 13.2. Carrier-phase receiver measurements

Although the required phase measurements are simple in concept, there are a number of practical difficulties. A typical laboratory frequency standard produces outputs at frequencies of 5 or 10 MHz; some masers also have an output at 100 MHz. The frequency of the GPS carrier, on the other hand, is about 1.5 GHz. It is very difficult to construct a phase meter that can operate at the carrier frequency directly, and so the usual arrangement is to mix the carrier with a local oscillator and to make the phase measurement at a lower intermediate frequency. The phase offset resulting from this mixing process is usually not important in a geodetic solution or in an application involving a frequency comparison (provided it does not change during the observing session). However, the offset must be determined for a time comparison. One way of insuring that this offset will be a constant is to derive the local oscillator signal that drives the mixer in the receiver from the external oscillator that is being measured, and some commercial receivers do this. A receiver that does not do this may report a step in the measured phase difference if its operation is interrupted (because of a power failure, for example).

<sup>4</sup> See also the many references to other carrier-phase experiments at the end of this paper.

The frequency of the carrier as seen by the receiver varies as the satellite moves through the sky because of changes in the Doppler shift and similar effects, and these effects must be removed by the receiver if it is to stay locked on the signal from the satellite. This is usually accomplished by mixing the output of the first mixer described above with a second local oscillator whose frequency and phase are adjusted by the tracking loop firmware in the receiver. If the receiver can track several satellites simultaneously, then there are several copies of this oscillator running in parallel, each having a different frequency and phase. These parallel oscillators are often synthesized using digital techniques to minimize the number of independent components that are needed to track several satellites.

Since all GPS satellites transmit using the same pair of carrier frequencies, the tracking loop of each channel in the receiver must utilize input from its own distinct code correlation processor to be sure that it remains locked on a single satellite. (This would not be necessary for a GLONASS receiver in principle, although it will be necessary in practice if more than one GLONASS satellite shares the same pair of carrier frequencies. Even if it were not necessary in principle, it would probably be useful to use the output of the code tracker to improve the noise performance of the carrier-phase measurement.)

The carrier phase estimates are usually obtained by integrating the error signal that is applied to the local oscillator that drives the second mixer stage. This error signal is usually a sum of contributions from both the code and carrier loop, with most of the weight assigned to the carrier loop. This second local oscillator is often realized using digital synthesis, so that the output datum would be the numerical value corresponding to the origin of the clock generator that drives the digital synthesizer. Any static phase offset in the first mixer stage is usually not included in this datum. In addition to the integer phase ambiguity in the model (equation (29)), an integration process has a constant which is related to the epoch at which the measurements were begun, and this constant will be different for each satellite. It cannot be determined by the tracking loop, and it is estimated during the post-processing phase of the analysis.

### 13.3. Carrier-phase analysis

The integrated Doppler offset reported by the receiver is compared to the predicted variation derived from a model of the motion of the satellite and the station. Since the wavelength of the carrier is about 19 cm, the variation predicted by the model must be at least this accurate, and a precise, post-processed ephemeris is usually required to do this. In addition to the motion of the satellite, the ephemeris contains estimates of the motion of the station with respect to the reference frame due to polar motion, and similar factors. A carrier-phase analysis comparing the clocks at two stations usually also includes observations from several additional stations. These additional data are very helpful in detecting cycle slips (a discontinuous change in the integer  $N$  in equation (29))—especially when the additional stations have very stable clocks and are near the primary stations of interest. This close proximity attenuates any inadequacies in the precise ephemerides and greatly facilitates the process of detecting cycle slips.

Since a cycle slip looks like a time step, our ability to detect it improves as the interval between data points is decreased. This is true for two reasons. In the first place, decreasing the interval between data points also decreases the time dispersion due to the frequency noise of the local oscillator. A slip of one cycle at  $L_1$ , for example, is equivalent to a time step of about 670 ps. If the interval between samples is 30 s (a typical value), then the frequency of the local oscillator must not vary by more than about  $2.2 \times 10^{-11}$  over this time interval if the step is to be identified with 50% probability. This tolerance would be correspondingly less stringent for more rapid data sampling. In the second place, more rapid data recording provides increased averaging of the white phase noise that characterizes the data at short time

intervals. Both of these considerations must be balanced against the increased storage and computing requirements needed to handle the additional data.

The precise orbits that are required for a carrier-phase analysis are available from the International GPS Service for Geodynamics (IGS). The IGS offers a number of different products with varying post-processing delays. At present, these delays are not a significant factor in carrier-phase comparisons between timing laboratories, which are not done in real time anyway. The IGS products and a description of the IGS analysis methods and services are available over the Internet from the IGS central bureau at: <http://igscb.jpl.nasa.gov>.

The additional delay due to the refractivity of the ionosphere can be estimated using the measured dispersion between the  $L_1$  and  $L_2$  carriers as described above. In normal circumstances, the  $L_2$  carrier is modulated by the encrypted P code. In spite of this fact, the dispersion between the two transmitted carriers can be estimated using the fact that both carriers have the same code, and all carrier-phase receivers operate in this ‘cross-correlation’ mode. In addition to the factor of three penalty in signal to noise ratio discussed above in connection with the ionospheric-free combination of the two signals, a multiple wavelength analysis has an additional complication in detecting cycle slips. This is because the phase step in the ‘ionospheric-free’ measurement (equation (13)) is no longer an exact integral value. The magnitude depends on which carrier frequency had the cycle slip, and there are obviously more complicated situations when both frequencies have simultaneous cycle slips.

The pseudo-range measurements have no measurement ambiguity, and the usual strategy is to combine carrier-phase data with code-based pseudo-range data so as to resolve the ambiguities in the former. Since the pseudo-range data have much lower resolution, a common strategy is to weight the pseudo-range data by a factor of 0.01 relative to the carrier-phase measurements. Although this strategy resolves the ambiguity in the carrier-phase data, it cannot be more accurate than the noise in the pseudo-range measurements themselves. Given this limitation, it is possible that carrier-phase analysis will offer no significant improvement over more conventional code-based receivers for time transfer applications.

Carrier-phase analyses are often performed on consecutive 24 h blocks of data, and the ambiguity resolution is computed independently in each block. Since the noise in the pseudo-range measurements enters into the ambiguity resolution algorithm, consecutive blocks generally have time offsets whose magnitudes are of the same order as the noise in the pseudo-range measurements averaged over 24 h. These apparent time steps are not important in geodetic applications, since such applications do not make use of the clock solution anyway. However, the absolute value of this delay is important in time distribution, and insuring that it remains constant is important in frequency measurements—especially when a frequency comparison spans a significant period of time (several days or longer). These are significant problems in practice, and it can be hard to remove a time step at a block boundary due to noise in the pseudo-range because of the time dispersion due to random-walk frequency noise of the clock. Even if the errors at each block boundary are randomly distributed with a zero mean, they introduce a random-walk time step at each block boundary which degrades the robustness of a multi-block frequency estimate.

#### **14. Two-way satellite time transfer**

This method is fundamentally different from the methods described above, all of which use passive receivers listening to the signals generated by and transmitted from a constellation of satellites, each of which contains an on-board frequency standard. The orbital periods of the satellites we have been discussing are approximately 12 h, so that many satellites are required to provide continuous global coverage.

Two-way satellite methods, on the other hand, use active transmitters on the ground, which transmit data to a transponder on a geostationary satellite (a satellite whose angular speed in its orbit exactly matches the angular speed of the Earth, so that it appears to remain in the same position relative to an observer on the Earth). A number of different satellites can be used for this purpose; the details depend on which one is chosen, but the general method is the same for all of them. The satellites are usually used for re-transmitting television broadcasts and similar information, which are transmitted up to the satellite from ground stations in real time.

Pulses from a local reference clock at 1 Hz are used to drive a PRN code generator. The system which has been used at NIST for some time uses eight selectable codes which are all  $10^4$  chips long. The clock in the code generator runs at a frequency of 2.5 MHz, so that the period of each chip is 400 ns. The entire code, therefore, requires 4 ms to transmit, so that there are 250 codes transmitted per second. The clock in the code generator is coherent with the 1 Hz ticks, and each tick synchronously inverts the sign of the PRN code. Although the numerical details are different, note that the modulation system is essentially identical to the scheme used by the GPS: the C/A code of that system has a slower chipping rate of 1.023 MHz but a faster repetition period of 1000 codes per second. The code that corresponds to the rollover to a new second is also identified by a data bit that reverses the sign of the C/A code in the GPS (see the previous discussion of the TLM and HOW in the GPS data stream).

The modems modulate a 70 MHz carrier with these PRN codes, so that there are 28 cycles of this carrier in each PRN chip. The carrier is up-converted to a frequency of about 14 GHz by the transmitting electronics. The 14 GHz signal is transmitted to a transponder in the satellite, where its modulation is re-transmitted on a second carrier with a frequency of about 12 GHz (this frequency range is called the Ku band). The 12 GHz signal is received at the remote end, where it is down-converted to 70 MHz. The 70 MHz carrier is passed to the modem where the PRN code is extracted.

The process of extracting the PRN code is conceptually the same as in the GPS. The receiving modem generates a local copy of the code, and the correlation between the local copy and the received signal is computed. The correlation process typically compares the received signal with two copies of the code: one is 0.5 chips early and the second is 0.5 chips late. The phase of the local oscillator is adjusted until the outputs of the two correlation processes are equal; the frequency of the local oscillator is kept locked to the incoming signal by maximizing the sum of the outputs of these two correlators. The correlation process is simpler than in the GPS case, since there is no need to adjust the frequency of the clock in this system once the modem has locked onto the received signal. (A GPS correlator must be ready to cope with a varying frequency caused by first-order variation in the Doppler shift.)

A conventional time interval counter at each station measures the time difference between the output tick from the modem and the corresponding pulses from the local clock. A measurement session usually lasts several minutes and therefore contains several hundred 1 s measurements. The measurements at each station are then exchanged after the fact using any convenient technique.

The analysis of the measurements to extract the time difference between the clocks at the two stations is straightforward. Suppose that a 1 s tick corresponding to some epoch occurs at stations A and B at times  $t_A$  and  $t_B$ , respectively. The goal in a time-transfer experiment would be to measure  $t_A - t_B$ ; if the goal is to compare the frequencies of the clocks at the two stations, then the time evolution of this difference (rather than its value at some epoch) would be the desired result.

If the one-way delay from A to B is  $d_{AB}$ , then the pulse that leaves station A at time  $t_A$  arrives at station B at time  $t_A + d_{AB}$ , and the time interval counter at station B, connected to

measure local—remote, will measure  $T_B = t_B - (t_A + d_{AB})$ . Most systems are full duplex and can transmit simultaneously in both directions. Therefore, while this measurement is in progress, an identical signal is going the other way. The time interval counter at station A is measuring  $T_A = t_A - (t_B + d_{BA})$ , where  $d_{BA}$  is the reverse one-way delay from B to A. Note that these one-way delays are composite values which include the delays through the station hardware, the troposphere, the ionosphere and the transponder in the satellite.

The delay through the transmitter section of the modem is the sum of two contributions: the actual delay through the hardware and the phase relationship between the 1 Hz ticks and the chipping clock, since the tick cannot have any effect until the start of the next chip (at the earliest). Since the two signals have a stable phase relationship to each other, the second contribution to the phase delay can be any constant value up to 400 ns (the period of one chip). The hardware delay is likely to be smaller but also less stable, since it will probably depend on the ambient temperature and similar effects.

The receiver will detect the inversion in the code that marks a tick as part of the correlation process. The number of chips that are required to detect this inversion varies with the signal to noise ratio of the measurement. A typical value might be a few hundred chips. These delays must be measured using loop-back and similar calibration procedures.

The analysis partitions the delays into two parts: a symmetric part that is the same in both directions, and an asymmetric part that is not. Apart from any hardware asymmetries (which are measured using ancillary measurements and then removed from the data), the most important contribution to the asymmetric part is due to the fact that the stations and the satellite are moving during the measurements (the Sagnac effect). The signal travelling West to East is in the same direction as this motion while the reverse direction is opposite to it. This asymmetry due to the Sagnac effect can be calculated from the locations of the stations and the position of the satellite, and this calculation need only be done once for any satellite/station path.

The analysis subtracts the two measurements  $T_A - T_B$ , to obtain  $\Delta T = T_A - T_B = 2(t_A - t_B) + 2(d_{AB} - d_{BA})$ . The symmetrical part of the delay cancels, and the asymmetries due to the hardware delays and the Sagnac effect are removed as described above. The magnitude of the symmetric path delay is not important; the method depends only on the fact that the delay must be the same in both directions. The uncertainty (and any time variation) of the calibration of the delay through the hardware at each station (and any asymmetry in the satellite transponders) will limit the accuracy of the method. Although all ground stations have some kind of calibration procedure, the station is a complex system, and it is very difficult to eliminate temperature-dependent effects. There is no way of evaluating the symmetry of the transponder in the satellite; this is not a problem in some satellites, which use the same transponder for transmissions in both directions.

The resolution of the method is a function of the chipping frequency, and newer modems that use a 20 MHz chipping frequency are being tested at NIST. In principle, this should increase the resolution (but not necessarily the accuracy) by a factor of 8. The resolution would also be increased if the number of ticks per second was increased, since the current system does not make optimum use of the 2.5 MHz chipping frequency—most of the 250 codes that are transmitted each second are not used.

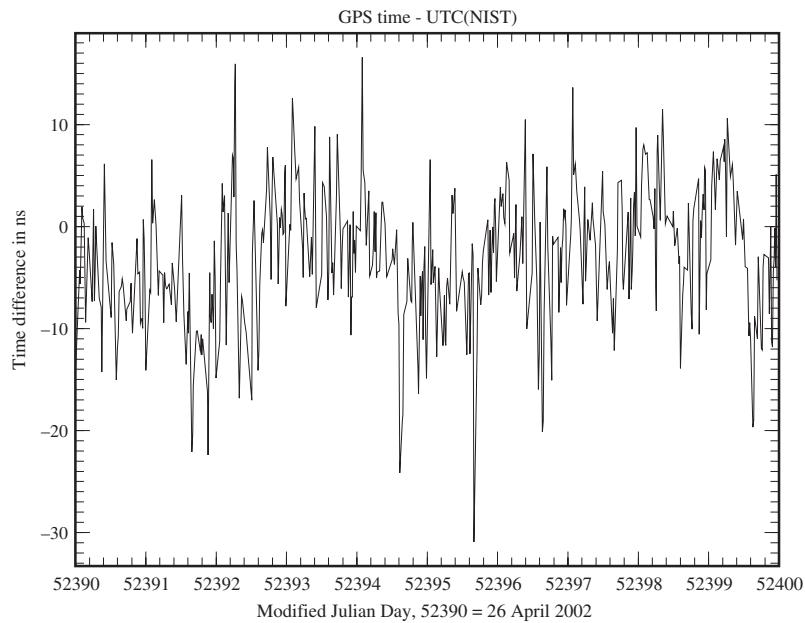
Since the clock that drives the PRN generator is derived from the same system that provides the 1 Hz ticks, it is also possible in principle to exploit this commonality using the equivalent of a GPS carrier-phase analysis on the phase of the 70 MHz carrier. As in the GPS analogue, this method is likely to be most useful for frequency comparisons, since the time tags that are an essential part of time difference measurements must still be estimated using the lower-frequency code transitions.

## 15. Time transfer data

In this section I will illustrate the principles of the preceding discussion using typical data. The first group of figures illustrate some of the aspects of time transfer using the GPS satellites that we have discussed. All of these figures use data from the same 10 day period—these data were chosen because they were the most recent data available when the manuscript was being prepared. On each plot the individual data points are connected by simple straight lines which are not otherwise significant. Unless otherwise noted in the figures, the data represent 13 min tracks computed as specified in the BIPM technical directives. No additional averaging or smoothing of the individual tracks has been applied and no tracks have been dropped.

Figure 3 illustrates the use of the GPS in a one-way code-based mode. This data set shows the output from a single-channel receiver, which can track only one satellite at a time. (This receiver was originally designed at the US National Bureau of Standards (NBS) about 20 yr ago. In addition to the receivers constructed by NBS, a nominally identical commercial version also exists. These receivers are widely used by many timing laboratories and NMIs.) The receiver measures the time difference between a local clock and GPS time once per second and then averages these data for 13 min using the BIPM tracking schedule and averaging method as discussed in the text. The reference clock for these data is UTC(NIST), a time signal derived from a weighted average of an ensemble of caesium standards and hydrogen masers located at the NIST laboratory in Boulder, Colorado. The details of the realization of UTC(NIST) are not important for this discussion, except that almost all of the variance in the figure is GPS measurement noise of one sort or another, and the contribution of noise in the local timescale is very small on the scale of the plot.

The very short-term noise in this figure (i.e. the variation between adjacent tracks) is approximately white phase noise with an amplitude of a few ns, and it represents the noise in

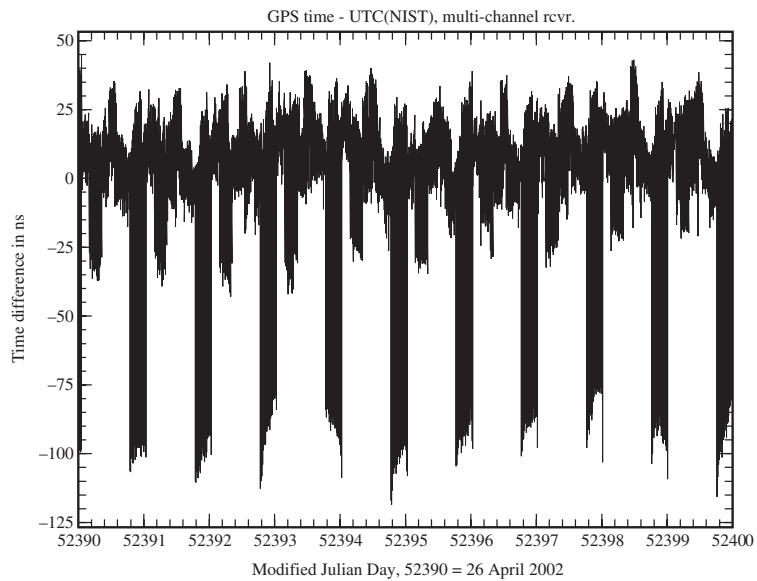


**Figure 3.** GPS time—UTC(NIST) measured using a single channel receiver. The time difference is measured every second, and these measurements are averaged as specified in the BIPM technical directives to produce a single value for each 13 min track.

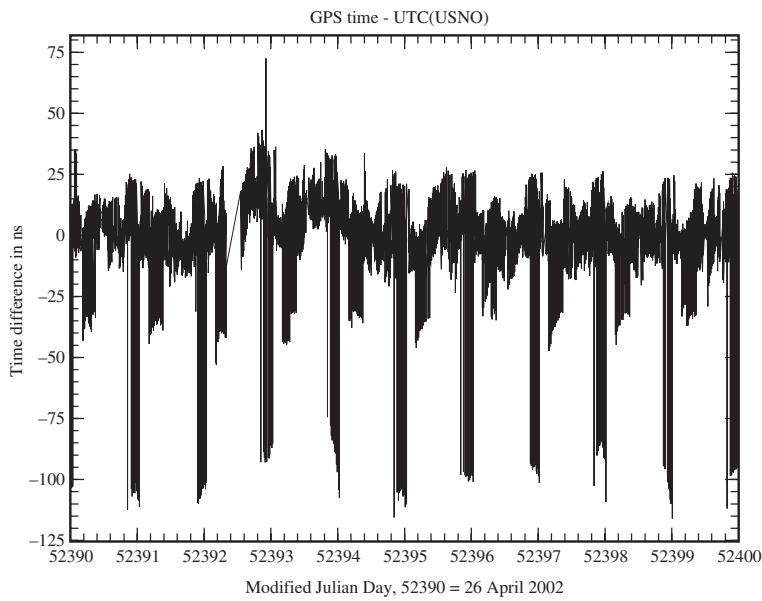
the receiver and the time interval counter that is used to measure the time differences between the local 1 Hz ticks and the corresponding ticks derived from the GPS signal. The longer-term structure in these data is a combination of contributions due to errors in the broadcast ephemeris (which is used in the receiver to construct the pseudo-range), errors in the model of the satellite clock, the inadequacy of the model of the ionosphere, the lack of any model of the troposphere, multi-path reflections from objects near the antenna, a variation in the response of the receiver due to changes in ambient temperature and probably other things as well. Most of the observed variance in this figure does not have a visible long-term component, and it can, therefore, be averaged to reduce the variance. Since the longer-term structure of the data are quite clearly not characteristic of a white noise process, an average is unlikely to be statistically robust or stationary. That is, both the details of the averaging method and the subset of the data that are used to construct the average will affect the result. In addition to these constraints, the local clock must have enough stability to support whatever averaging process is used. For example, a very good quality commercial caesium standard has a time dispersion of about  $2 \text{ ns day}^{-1}$ , so that a clock of this caliber would be very useful in removing almost all of the observed variance by an appropriate averaging scheme whose time constant was on the order of a few days. Another way of saying this is that the free-running stability of such a clock is better than the signal received from the GPS for shorter averaging times.

A comparison between the free-running stability of some other clock and the variance of the GPS data can be used to construct an optimum averaging algorithm for any other device. Although the details will vary with the particular situation, most atomic frequency standards will be more stable than the measured GPS time differences at sufficiently short times, so that some averaging will always be appropriate.

The data in figure 3 represent the 13 min tracks specified in the BIPM tracking schedule when only a single satellite is tracked at any time. These are only a small fraction of the data that could be acquired during the same time period. Figures 4 and 5 show the data from



**Figure 4.** GPS time—UTC(NIST) measured using a multi-channel receiver than can track up to eight satellites simultaneously. The data are acquired and averaged as specified in the BIPM technical directives.

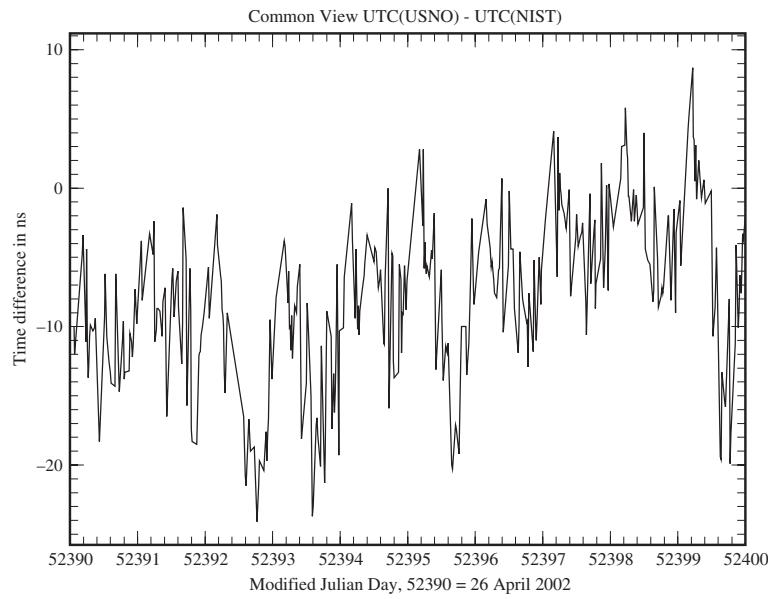


**Figure 5.** GPS time—UTC(USNO) measured using a multi-channel receiver that can track up to eight satellites simultaneously. The data are acquired and averaged as specified in the BIPM technical directives.

multi-channel receivers for the same time period. Each point represents the same 13 min average, but these receivers can track up to eight satellites simultaneously, and it reports data from all satellites in view. Figure 4 shows the data from a receiver at NIST and figure 5 shows the data from an essentially identical receiver at the US Naval Observatory in Washington. In both cases, the reference clock for the measurements is a local realization of UTC; the noise in both local timescales can be ignored on the scale of these plots.

Although both of these figures show many more tracks, the data are not statistically better than those in figure 3. There are periods when the short-term variance is about the same as on figure 3, but there are also periods where the noise is much greater. These noisy periods are usually caused by the receiver reporting data from satellites that have low elevation angles so that the effect of multi-path is larger, the direct signal is weaker and the contributions of both the troposphere and the ionosphere are largest. However, even in the best of times, the variance is not much better than the variance for the data in figure 3. This should not be too surprising—the contributors to the variance are largely due to slowly varying systematic sources and not white phase noise or other random processes with similar power spectra that are amenable to averaging.

Figure 6 shows the common-view time difference between USNO and NIST. It is constructed using the data sets shown in figures 3 and 5. The corresponding tracks for each satellite at each epoch are subtracted point by point. Since the data in figure 3 are from a single channel receiver, it is this data set that really defines the number of common-view tracks that are available. As we discussed above, a common-view difference cancels any effects that are highly correlated in the two data sets, and the improvement is quite noticeable. The short-term noise in figure 6 is from sources that are not well correlated between the two stations. These include some portion of the contribution of the ionosphere, the effect of the troposphere, multi-path effects at the two receivers and any sensitivity of the receivers to fluctuations in the ambient temperature. In spite of these residual noise sources, the common-view cancellation



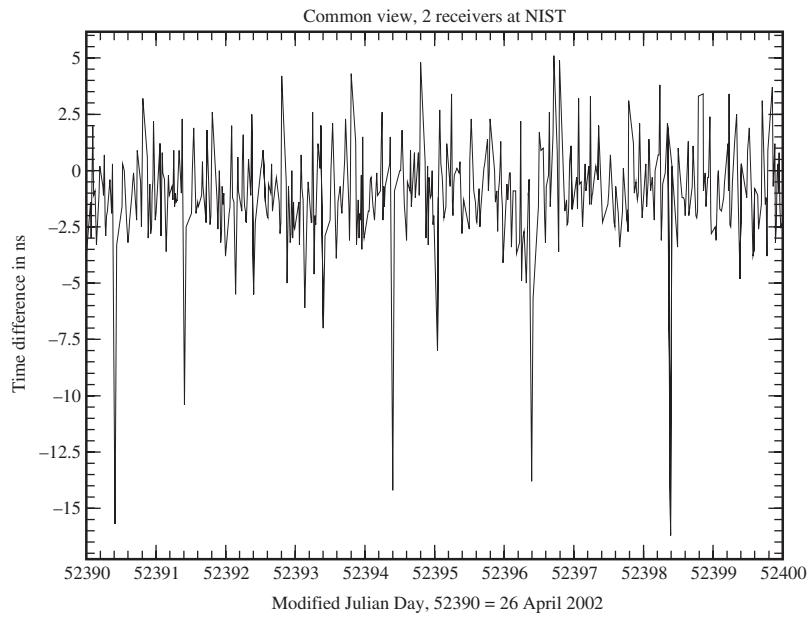
**Figure 6.** Common-view difference between UTC(NIST) and UTC(USNO). These data are the point by point differences between corresponding points in figures 3 and 5. (A corresponding point is a 13 min track of the same satellite at the same time at both stations.)

is good enough so that it is now possible to see the frequency difference between UTC(USNO) and UTC(NIST). The fractional frequency difference between the two timescales is usually less than  $1 \times 10^{-14}$ , which implies a slowly varying time dispersion on the order of a few ns per day. (Both timescales are independently steered to UTC as defined by the BIPM so that there is no long-term divergence between them.)

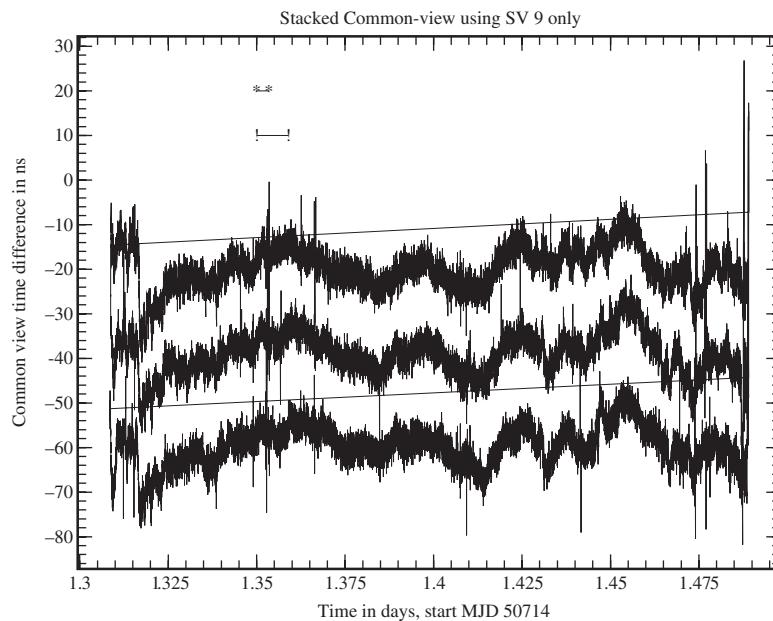
Figure 7 displays the best that common view can provide. The data on this plot show the common-view time differences between two identical receivers at NIST that are referenced to the same clock. The antennae are a short distance apart on the roof of NIST, so that the contributions of the ionosphere and the troposphere are essentially identical and cancel in the differences as a result. The residual variance is due to noise in the receiver, the effects of fluctuations in the ambient temperature that are not identical and the effects of multi-path reflections. The largest quasi-periodic outliers are probably due to differences in the multi-path.

The BIPM tracking schedule advances the time of a track by 4 min every day, so that the geometrical relationship between the satellite and the receiving antennae is almost unchanged from one day to the next. Since the effects of multi-path reflections depend on the geometrical relationship between the receiving antenna, nearby reflectors and the line of sight to the satellite, these reflections tend to be the same from day to day for each satellite that is tracked. This has the unfortunate side effect of making it difficult to evaluate the contribution of these multi-path reflections, since the algorithm used to define the tracking schedule converts them to nearly constant offsets. Although these offsets obviously vary from satellite to satellite since each one is at a different elevation and azimuth during a track, it is usually impossible to separate the contribution of multi-path from other satellite-dependent errors, such as those that result from orbit errors in the broadcast ephemeris or from errors in the assumed position of the receiving antenna.

Figure 8 illustrates the problem. The data in this figure show the short-baseline common-view time differences between two receivers at NIST. The raw 1 s time difference measurements



**Figure 7.** Short-baseline common view between two identical receivers at NIST. The two receivers measure the difference between GPS time and UTC(NIST), and the common-view data are computed point by point as in the previous figure.



**Figure 8.** Short baseline common view between two identical receivers at NIST. The plot shows the differences between the raw, 1 s data with no averaging. Data from successive data are plotted with a time shift of  $-4$  min and are offset vertically for clarity.

are shown with no averaging. The various traces in the figure are ‘stacked’, i.e. they show the time differences obtained on consecutive days. Each plot is advanced in time by 4 min, and each curve is offset vertically by an arbitrary amount for clarity. Note the very high correlation in the variance from day to day. Also note how the character of the variance changes as the satellite elevation decreases towards the end of the time period. As we would expect, the multi-path reflections get larger and more rapid as the elevation of the satellite decreases. These data show the differential effect between two receivers whose antennae are about 2 m apart. It is not possible to compute the actual contribution of multi-path reflections to each station, but it is reasonable to expect that the contribution is several ns on average and might be as large as 15 ns under some circumstances.

The horizontal lines on the plot shown as (\*—\*) and (!—!) show time intervals of 5 min, and 13 min, respectively. Both of these time are commonly used for averaging GPS data. (The 13 min interval is the one specified by the BIPM.) The multi-path variations displayed in the figure are not characterized by white phase noise over either of these averaging intervals, so that an averaging process using either of these time intervals is not statistically robust. Although any averaging process over these time intervals will produce a value, the value will depend on exactly which portion of the data are used and how the average is computed. The bias implicit in this value will be nearly constant from day to day for any one satellite at any one station, but it is generally not attenuated in common view because it is a strong function of the locations of the reflectors at each site. However, any change in the averaging algorithm (or combining data acquired using different algorithms) is likely to introduce a change in the measured time difference because of the complex interaction between the details of the algorithm and the time dependence of the multi-path reflections.

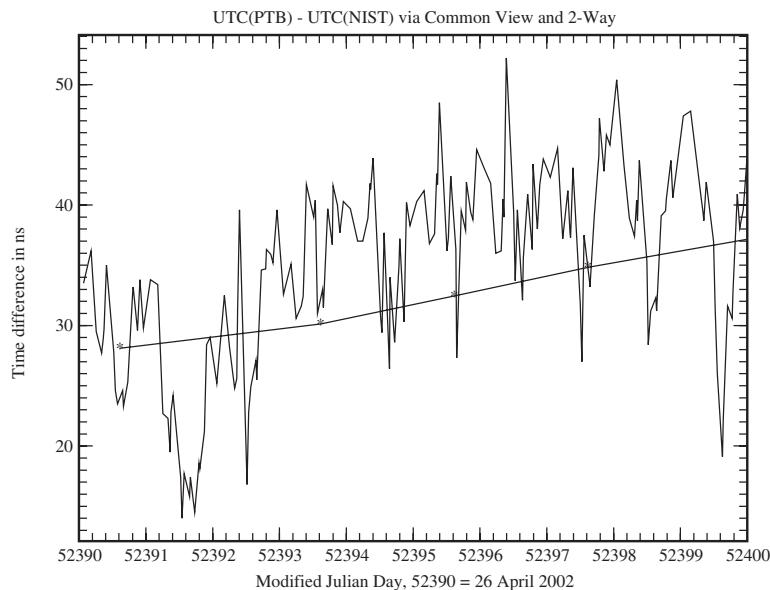
This argument suggests that the details of any algorithm used to average the 1 s time difference measurements are less important than the fact that the same algorithm is used every day at all sites. Since the data are characterized by white phase noise only for short averaging times, this argument would suggest that the 13 min track time currently specified by the BIPM is too long, and that more frequent and shorter tracks would be a better choice. Furthermore, the rather complex least-squares fitting procedures that are part of the BIPM technical specifications would seem to be besides the point—the data show in figure 8 are simply not well modelled by these linear and quadratic functions of time, so that while the slopes and intercepts that result from the BIPM averaging algorithm are well defined and numerically stable, they have little physical significance. Simpler models and shorter tracks would produce values that were different because they averaged the non-stationary variations in a different way, but these models would not be statistically worse in any quantifiable way.

We would expect that the common-view method would tend to be less effective as the distance between the two receivers is increased because of the corresponding decrease in the correlation between the two path delays. Figure 9 shows the common-view time differences between UTC(NIST) and the timescale UTC(PTB), which is maintained by the Physikalisch-Technische Bundesanstalt in Braunschweig, Germany. The baseline between these two sites is about one-third of the circumference of the Earth—about 4 times longer than the distance between NIST in Colorado and the USNO in Washington. The short-term noise in the figure is less apparent because the longer baseline means that there will be fewer satellites in common view so that the mean interval between tracks is longer. As expected, the diurnal variations are larger because these variations are less well correlated at this distance. The long period variation is caused by the varying frequency offset between UTC(NIST) and UTC(PTB). This frequency offset produces time dispersions on the order of ns per day; as with the USNO, UTC(PTB) is steered to UTC as defined by the BIPM so that there is no long-term divergence between the timescales maintained at the different laboratories.

Figure 9 also shows the time difference between UTC(PTB) and UTC(NIST) as measured using two-way satellite time transfer. These measurements are made only 3 times per week (usually on Monday, Wednesday and Friday) for 2 min, and there are five measurements (identified with the symbol (\*)) during the period shown in the figure. The measurements are made using the Intelsat 706 satellite, which is located on the Equator near longitude 53° W. This satellite uses different transponders for signals going from East to West and from West to East, and this difference might limit the accuracy of the time transfer, since the delays through the transponders might not be identical. The measurements shown here use the Ku band frequencies (14 GHz up-link, 12 GHz down-link). The same satellite also supports lower-frequency communication in the C band (6 GHz up-link, 4 GHz down-link). The general method would be the same using the lower frequency, although larger antennae would be required to realize the same signal to noise ratio.

It is clear from this figure that two-way satellite time transfer has much less noise than GPS code-based measurements, even with the advantages of common-view subtraction. Two-way links are, therefore, replacing common-view GPS in international time and frequency coordination, and in other applications requiring very good precision, such as comparing primary frequency standards.

The current generation of primary frequency standards realize the SI second with an uncertainty of about  $2 \times 10^{-15}$ , and primary standards with even better performance are on the horizon. This level of accuracy implies a time dispersion on the order of  $0.2 \text{ ns day}^{-1}$ , and it is clear that the noise in code-based GPS is too large to permit two such standards at different laboratories to be compared in a reasonable averaging time. Such comparisons are important in practice because data from these primary frequency standards are used to define the length of the SI second. Comparing different standards, especially those that use different



**Figure 9.** Difference between UTC(PTB) and UTC(NIST) measured using common-view GPS and two-way satellite time transfer. The common-view data is acquired and processed as in the previous figures. The two-way satellite time data are measured using three sessions per week. Each session consists of hundred and twenty 1 s measurements using a Ku band system as described in the text.

techniques, is also useful as a check on the different implementations of the realization of the standard frequency.

In addition to two-way satellite time transfer, carrier-phase GPS also has a resolution that would be useful in primary frequency standard comparisons. Since the two methods are very different, simultaneous comparisons between the two methods operating between the same two stations would be very useful. A number of such experiments have been performed, and additional ones are currently in progress. One ongoing experiment is a comparison of the primary frequency standards at PTB and NIST using both two-way satellite time transfer and carrier-phase GPS simultaneously. The first results of this comparison [18] showed agreement between the two methods to better than  $2 \times 10^{-15}$ , which is less than the combined uncertainties of the standards in the two laboratories. The contribution of the noise in the time transfer process itself was about  $4.2 \times 10^{-16}$ , which is somewhat less than 30% of the total uncertainty. This value is the combined uncertainty in the two methods and it is not possible to know what fraction of this uncertainty is due to each of the methods. The time difference between the two methods has a possible secular trend of about  $0.03 \text{ ns day}^{-1}$ , which corresponds to a fractional frequency offset of about  $3.5 \times 10^{-16}$ . The source of this trend is not known, but it is possible that it arises from the details of the carrier-phase processing, and we are continuing to work on improving this.

## 16. Conclusions

In this paper I have outlined the principles of distributing time and frequency information using navigation satellites (such as GPS) and communications satellites. Both of these types of satellites are currently used both for international time and frequency coordination and in more routine time and frequency distribution systems. Although it is difficult to provide a single estimate of the performance of these systems in all of their different configurations, it is generally true that it is easy to use them to transmit time with an uncertainty of  $1 \mu\text{s}$  and it is very difficult (perhaps impossible at the present time) to transmit time with an uncertainty of  $1 \text{ ns}$ .

Since frequency transfer is generally implemented as the difference between two time transfer measurements separated by some averaging time interval, the corresponding capabilities with respect to frequency transfer depend only on the stability of the various parameters (such as the path delay) and not on their absolute values. This stability requirement is somewhat weaker than the accuracy requirement needed for time transfer applications. For example, the contribution of the communication channel to the uncertainty in the comparison of the primary frequency standards at NIST and PTB was  $4.2 \times 10^{-16}$ . This is equivalent to a time deviation ( $\sigma_x$ ) of about  $0.2 \text{ ns}$  over the averaging interval of about 3 weeks. It would be very difficult to realize absolute time transfer at this level of uncertainty.

All of the time and frequency distribution methods I have discussed depend on using a satellite system which was primarily designed for another purpose. Two-way satellite time transfer, for example, uses communication satellites whose primary function is to transmit television signals and similar information. This principal application does not require sub-nanosecond stability in the symmetry of the delays through the station hardware or the satellite transponders, and the equipment is, therefore, not primarily designed to satisfy this requirement. Likewise, carrier-phase GPS measurements are often realized using receivers and analysis software that were primarily designed for geodetic applications, where the delay through the receiver and the accuracy of the local clock solution are not very important.

Even multi-path reflections, which would seem to be just as serious a problem for geodesy as for time transfer, often have less of an impact on the former than on the latter. This is because the statistics of the clocks at the receiver stations are much less favourable than the statistics of the position of the antenna, so that geodetic determinations can support much

longer averaging times than clock comparisons between the same sites. The effects of multi-path are both quasi-periodic and bounded, and can therefore be attenuated by averaging times that span several days. This is straightforward for geodetic solutions, but would be difficult or impossibly long for general time transfer because of the flicker and random-walk noise processes in many of the station clocks.

In spite of these limitations, both time and frequency comparisons are often characterized by white phase noise at sufficiently short-time intervals, so that some amount of averaging is almost always appropriate. The exact averaging interval depends on the stability of the clock in the receiver and on the spectrum of the fluctuations in the delay along the path back to the transmitter. Typical averaging intervals might range from a few seconds for a receiver with a simple quartz-crystal oscillator to several days for a caesium standard. The long averaging time that was used in the carrier-phase comparison of primary frequency standards described above was a special case, since such standards are presumed to be characterized by white frequency noise for all averaging times, an assumption that is very definitely not true for commercial atomic frequency standards.

The most promising new techniques are carrier-phase GPS and two-way satellite time transfer. As we discussed above, both of these techniques have been used to compare primary frequency standards with a fractional frequency uncertainty in the transfer process itself of less than  $1 \times 10^{-15}$ . The averaging times required to realize these comparisons were several weeks long, and future work in the field will be aimed at improving the short-term noise of the comparisons so that the same level of uncertainty can be realized with much shorter averaging times. These improvements will almost certainly be needed in order to support a new generation of frequency standards which will have fractional frequency uncertainties of significantly less than  $1 \times 10^{-15}$ .

Finally, there is the question of the traceability of satellite time signals to national and international standards. (By traceability I mean that there is an unbroken chain of measurements between the signal received by an end user and a national or international time standard, such as a NMI or timing laboratory. Each of these measurements must be characterized by its uncertainty.)

Using the methods that we have described above, it is clear that signals from any of the navigation satellites can be made traceable to national standards. The GPS time is steered towards UTC(USNO), and signals from the GPS satellites contain a prediction of any residual time difference between these two timescales; the signals from GLONASS and Galileo contain (or will contain) similar information. Although the parameters in the GPS message are technically predictions rather than measurements, this is a technical distinction without much of a practical difference for most users, since the errors in the predictions are only on the order of tens of ns or less. Users who need the greater accuracy that can be realized using actual measurements rather than these predictions can use common-view methods. Many laboratories, including the US Naval Observatory and NIST publish GPS tracking data (typically with a short delay of 1 day or less) to support these users. These data would allow a user to realize direct traceability to many timing laboratories and NMIs with an accuracy that was independent of the performance of the satellite system in first order. As a practical matter, using any of these satellite systems in common view provides general users with the highest-accuracy access to the real-time UTC timescales of most timing laboratories and NMIs.

For most users, the weakest link in the traceability chain is the calibration of the receiving equipment, including an estimate of the effects of multi-path at the user's site. These effects are usually not important until the required uncertainty becomes less than  $1 \mu\text{s}$ , and they usually dominate all other sources of uncertainty when the accuracy requirement is on the order of ns. They are not improved by using common-view techniques.

Users who need legal (in addition to purely technical) traceability may face an additional hurdle in that just doing the ‘right thing’ may not be adequate [19]. They may also have to be able to prove that they did the ‘right thing’ in some future judicial proceeding. Although there are few if any precedents in this area, it is possible that a simple broadcast-only system may never be adequate in this case, and that an additional certification by a disinterested third party may also be required [20]. This certification may require an additional communication channel that cannot be realized by any broadcast type of time distribution system, whether it is based on transmissions from satellites or from terrestrial sources.

## References

- [1] Levine J 1999 Introduction to time and frequency metrology *Rev. Sci. Instrum.* **70** 2567–96
- [2] Sullivan D B, Allan D W, Howe D A and Walls F L 1990 Characterization of clocks and oscillators *NIST Technical Note* 1337. Published by NIST and available from the Time and Frequency Division, National Institute of Standards and Technology, Boulder, Colorado 80305, USA
- [3] www.bipm.org
- [4] US Department of Defense World Geodetic System 1984 Its definition and relationships with local geodetic systems *National Imagery and Mapping Agency (Bethesda, Maryland, 1997)* Document TR8350.2
- [5] Boykov V V, Galazin V F, Kaplan B L, Maximov V G and Bazlov Yu A 1993 Experimental of compiling the geocentric system of coordinates PZ-90, *Geodeziya i Kartografiya*, pp 18–21
- [6] McCarthy D D (ed) 1996 IERS Conventions *IERS Technical Note* 21, Paris, Observatoire de Paris
- [7] Jorgensen P S 1989 An assessment of ionospheric effects on the user *J. Inst. Navigation* **36** 8–14
- [8] Smith E and Weintraub S 1953 The constants in the equation for atmospheric refractive index at radio frequencies *Proc. IRE* **41** 1037–53
- [9] Resch G M 1980 Water vapor—the wet blanket of microwave interferometry *Proc. Intl. Workshop on Atmospheric Water Vapor (11–13 September 1979)* ed A Deepak (New York: Academic)
- [10] Gary B L, Keihm S J and Janssen M A 1985 Optimum strategies and performance for the remote sensing of path delay using ground-based microwave radiometers *IEEE Trans. Geosci. Remote Sensing* **GE-23** 479–84
- [11] Walter S J and Bender P L 1992 The slant path atmospheric refraction calibrator: an instrument to measure the microwave propagation delays induced by atmospheric water vapor *IEEE Trans. Geosci. Remote Sensing* **GE-30** 462–71  
Walter S J 1990 *PhD Thesis* Measurement of the microwave propagation delay induced by atmospheric water vapor, University of Colorado, Boulder
- [12] Owens J C 1967 Optical refractive index of air: dependence on pressure, temperature and composition *Appl. Opt.* **6** 51–8
- [13] Edlen B 1966 The refractive index of air *Metrologia* **2** 71–80
- [14] Gold R 1967 Optimal binary sequences for spread spectrum multiplexing *IEEE Trans. Info. Theory* **33** 619–21
- [15] Allan D W and Thomas C 1994 Technical directives for standardization of GPS time receiver software *Metrologia* **31** 69–79
- [16] Levine J 1999 Time transfer using multi-channel GPS receivers *IEEE Trans. UFFC* **46** 392–8
- [17] Larson K M and Levine J 1999 Carrier-phase time transfer *IEEE Trans. UFFC* **46** 1001–12
- [18] Parker T, Hetzel P, Jefferts S, Weyers S, Nelson L, Bauch A and Levine J 2001 First comparison of remote cesium fountains *Proc. 2001 Frequency Control Symposium (Seattle, Washington)* pp 63–8 *IEEE Catalog Number* 01CH37218, ISBN: 0-7803-7028-7
- [19] Levine J 2001 GPS and the legal traceability of time *GPS World* pp 52–8
- [20] Levine J 2001 Legal traceability of time *Symp. National Conf. of Standards Laboratories (NCSL) (Washington DC, July 2001)* (Boulder, Colorado: NCSL International)

## Bibliography

In addition to the references listed above, the following references provide additional general information on the satellite systems I have discussed.

- Kaplan E D (ed) 1996 *Understanding GPS Principles and Applications* (Artech House Publishers, Boston) ISBN: 0-89006-793-7
- Alfred Leick (ed) 1990 *GPS Satellite Surveying* (New York: Wiley) ISBN: 0-471-81990-5
- The Global Positioning System *Papers published in Navigation* (Washington, DC: Institute of Navigation) ISBN: 0-936406-01-1