

Relativity in the Global Positioning System

Neil Ashby

Dept. of Physics, University of Colorado

Boulder, CO 80309-0390

U.S.A.

Neil.Ashby@mobek.colorado.edu

http://physics.colorado.edu/faculty/ashby_n.html

Published on 28 January 2003

www.livingreviews.org/Articles/Volume6/2003-1ashby

Living Reviews in Relativity

Published by the Max Planck Institute for Gravitational Physics

Albert Einstein Institute, Germany

Abstract

The Global Positioning System (GPS) uses accurate, stable atomic clocks in satellites and on the ground to provide world-wide position and time determination. These clocks have gravitational and motional frequency shifts which are so large that, without carefully accounting for numerous relativistic effects, the system would not work. This paper discusses the conceptual basis, founded on special and general relativity, for navigation using GPS. Relativistic principles and effects which must be considered include the constancy of the speed of light, the equivalence principle, the Sagnac effect, time dilation, gravitational frequency shifts, and relativity of synchronization. Experimental tests of relativity obtained with a GPS receiver aboard the TOPEX/POSEIDON satellite will be discussed. Recently frequency jumps arising from satellite orbit adjustments have been identified as relativistic effects. These will be explained and some interesting applications of GPS will be discussed.

Article Amendments

On author request a *Living Reviews* article can be amended to include errata and small additions to ensure that the most accurate and up-to-date information possible is provided. For detailed documentation of amendments, please go to the article's online version at

<http://www.livingreviews.org/Articles/Volume6/2003-1ashby/>.

Owing to the fact that a *Living Reviews* article can evolve over time, we recommend to cite the article as follows:

Ashby, N.,
“Relativity in the Global Positioning System”,
Living Rev. Relativity, **6**, (2003), 1. [Online Article]: cited on <date>,
<http://www.livingreviews.org/Articles/Volume6/2003-1ashby/>.

The date in 'cited on <date>' then uniquely identifies the version of the article you are referring to.

Contents

1	Introduction	4
2	Reference Frames and the Sagnac Effect	7
3	GPS Coordinate Time and TAI	11
4	The Realization of Coordinate Time	15
5	Relativistic Effects on Satellite Clocks	17
6	TOPEX/POSEIDON Relativity Experiment	22
7	Doppler Effect	26
8	Crosslink Ranging	28
9	Frequency Shifts Induced by Orbit Changes	30
10	Secondary Relativistic Effects	40
11	Applications	42
12	Conclusions	43
	References	44

1 Introduction

The Global Positioning System (GPS) can be described in terms of three principal “segments”: a Space Segment, a Control Segment, and a User Segment. The Space Segment consists essentially of 24 satellites carrying atomic clocks. (Spare satellites and spare clocks in satellites exist.) There are four satellites in each of six orbital planes inclined at 55° with respect to earth’s equatorial plane, distributed so that from any point on the earth, four or more satellites are almost always above the local horizon. Tied to the clocks are timing signals that are transmitted from each satellite. These can be thought of as sequences of events in spacetime, characterized by positions and times of transmission. Associated with these events are messages specifying the transmission events’ spacetime coordinates; below I will discuss the system of reference in which these coordinates are given. Additional information contained in the messages includes an almanac for the entire satellite constellation, information about satellite vehicle health, and information from which Universal Coordinated Time as maintained by the U.S. Naval Observatory – UTC(USNO) – can be determined.

The Control Segment is comprised of a number of ground-based monitoring stations, which continually gather information from the satellites. These data are sent to a Master Control Station in Colorado Springs, CO, which analyzes the constellation and projects the satellite ephemerides and clock behaviour forward for the next few hours. This information is then uploaded into the satellites for retransmission to users.

The User Segment consists of all users who, by receiving signals transmitted from the satellites, are able to determine their position, velocity, and the time on their local clocks.

The GPS is a navigation and timing system that is operated by the United States Department of Defense (DoD), and therefore has a number of aspects to it that are classified. Several organizations monitor GPS signals independently and provide services from which satellite ephemerides and clock behavior can be obtained. Accuracies in the neighborhood of 5–10 cm are not unusual. Carrier phase measurements of the transmitted signals are commonly done to better than a millimeter.

GPS signals are received on earth at two carrier frequencies, L1 (154×10.23 MHz) and L2 (120×10.23 MHz). The L1 carrier is modulated by two types of pseudorandom noise codes, one at 1.023 MHz – called the Coarse/Acquisition or C/A-code – and an encrypted one at 10.23 MHz called the P-code. P-code receivers have access to both L1 and L2 frequencies and can correct for ionospheric delays, whereas civilian users only have access to the C/A-code. There are thus two levels of positioning service available in real time, the Precise Positioning Service utilizing P-code, and the Standard Positioning Service using only C/A-code. The DoD has the capability of dithering the transmitted signal frequencies and other signal characteristics, so that C/A-code users would be limited in positioning accuracy to about ± 100 meters. This is termed Selective Availability, or SA. SA was turned off by order of President Clinton in May, 2000.

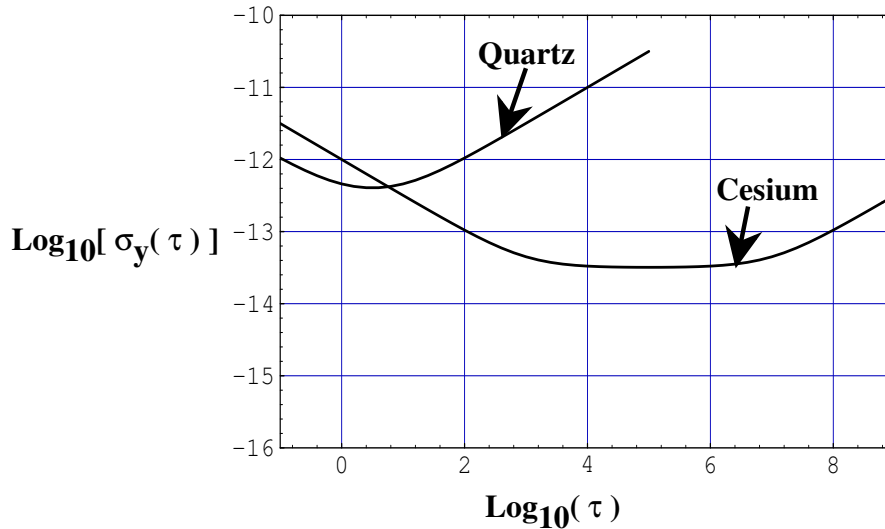


Figure 1: *Typical Allan deviations of Cesium clocks and quartz oscillators, plotted as a function of averaging time τ .*

The technological basis for GPS lies in extremely accurate, stable atomic clocks. Figure 1 gives a plot of the Allan deviation for a high-performance Cesium clock, as a function of sample time τ . If an ensemble of clocks is initially synchronized, then when compared to each other after a time τ , the Allan deviation provides a measure of the rms fractional frequency deviation among the clocks due to intrinsic noise processes in the clocks. Frequency offsets and frequency drifts are additional systematic effects which must be accounted for separately. Also on Figure 1 is an Allan deviation plot for a Quartz oscillator such as is typically found in a GPS receiver. Quartz oscillators usually have better short-term stability performance characteristics than Cesium clocks, but after 100 seconds or so, Cesium has far better performance. In actual clocks there is a wide range of variation around the nominal values plotted in Figure 1.

The plot for Cesium, however, characterizes the best orbiting clocks in the GPS system. What this means is that after initializing a Cesium clock, and leaving it alone for a day, it should be correct to within about 5 parts in 10^{14} , or 4 nanoseconds. Relativistic effects are huge compared to this.

The purpose of this article is to explain how relativistic effects are accounted for in the GPS. Although clock velocities are small and gravitational fields are weak near the earth, they give rise to significant relativistic effects. These effects include first- and second-order Doppler frequency shifts of clocks due to their relative motion, gravitational frequency shifts, and the Sagnac effect due to earth's rotation. If such effects are not accounted for properly, unacceptably large errors in GPS navigation and time transfer will result. In the GPS one can find many examples of the application of fundamental relativity principles.

These are worth careful study. Also, experimental tests of relativity can be performed with GPS, although generally speaking these are not at a level of precision any better than previously existing tests.

The principles of position determination and time transfer in the GPS can be very simply stated. Let there be four synchronized atomic clocks that transmit sharply defined pulses from the positions \mathbf{r}_j at times t_j , with $j = 1, 2, 3, 4$ an index labelling the different transmission events. Suppose that these four signals are received at position \mathbf{r} at one and the same instant t . Then, from the principle of the constancy of the speed of light,

$$c^2(t - t_j)^2 = |\mathbf{r} - \mathbf{r}_j|^2, \quad j = 1, 2, 3, 4. \quad (1)$$

where the defined value of c is exactly $299792458 \text{ m s}^{-1}$. These four equations can be solved for the unknown space-time coordinates $\{\mathbf{r}, t\}$ of the reception event. Hence, the principle of the constancy of c finds application as the fundamental concept on which the GPS is based. Timing errors of one ns will lead to positioning errors of the order of 30 cm. Also, obviously, it is necessary to specify carefully the reference frame in which the transmitter clocks are synchronized, so that Eq. (1) is valid.

The timing pulses in question can be thought of as places in the transmitted wave trains where there is a particular phase reversal of the circularly polarized electromagnetic signals. At such places the electromagnetic field tensor passes through zero and therefore provides relatively moving observers with sequences of events that they can agree on, at least in principle.

2 Reference Frames and the Sagnac Effect

Almost all users of GPS are at fixed locations on the rotating earth, or else are moving very slowly over earth's surface. This led to an early design decision to broadcast the satellite ephemerides in a model earth-centered, earth-fixed, reference frame (ECEF frame), in which the model earth rotates about a fixed axis with a defined rotation rate, $\omega_E = 7.2921151467 \times 10^{-5} \text{ rad s}^{-1}$. This reference frame is designated by the symbol WGS-84(G873) [19, 3]. For discussions of relativity, the particular choice of ECEF frame is immaterial. Also, the fact the earth truly rotates about a slightly different axis with a variable rotation rate has little consequence for relativity and I shall not go into this here. I shall simply regard the ECEF frame of GPS as closely related to, or determined by, the International Terrestrial Reference Frame established by the IERS.

It should be emphasized that the transmitted navigation messages provide the user only with a function from which the satellite position can be calculated *in the ECEF* as a function of the transmission time. Usually, the satellite transmission times t_j are unequal, so the coordinate system in which the satellite positions are specified changes orientation from one measurement to the next. Therefore, to implement Eqs. (1), the receiver must generally perform a different rotation for each measurement made, into some common inertial frame, so that Eqs. (1) apply. After solving the propagation delay equations, a final rotation must usually be performed into the ECEF to determine the receiver's position. This can become exceedingly complicated and confusing. A technical note [10] discusses these issues in considerable detail.

Although the ECEF frame is of primary interest for navigation, many physical processes (such as electromagnetic wave propagation) are simpler to describe in an inertial reference frame. Certainly, inertial reference frames are needed to express Eqs. (1), whereas it would lead to serious error to assert Eqs. (1) in the ECEF frame. A "Conventional Inertial Frame" is frequently discussed, whose origin coincides with earth's center of mass, which is in free fall with the earth in the gravitational fields of other solar system bodies, and whose z -axis coincides with the angular momentum axis of earth at the epoch J2000.0. Such a local inertial frame may be related by a transformation of coordinates to the so-called International Celestial Reference Frame (ICRF), an inertial frame defined by the coordinates of about 500 stellar radio sources. The center of this reference frame is the barycenter of the solar system.

In the ECEF frame used in the GPS, the unit of time is the SI second as realized by the clock ensemble of the U.S. Naval Observatory, and the unit of length is the SI meter. This is important in the GPS because it means that local observations using GPS are insensitive to effects on the scales of length and time measurements due to other solar system bodies, that are time-dependent.

Let us therefore consider the simplest instance of a transformation from an inertial frame, in which the space-time is Minkowskian, to a rotating frame of reference. Thus, ignoring gravitational potentials for the moment, the metric in

an inertial frame in cylindrical coordinates is

$$-ds^2 = -(c dt)^2 + dr^2 + r^2 d\phi^2 + dz^2, \quad (2)$$

and the transformation to a coordinate system $\{t', r', \phi', z'\}$ rotating at the uniform angular rate ω_E is

$$t = t', \quad r = r', \quad \phi = \phi' + \omega_E t', \quad z = z'. \quad (3)$$

This results in the following well-known metric (Langevin metric) in the rotating frame:

$$-ds^2 = -\left(1 - \frac{\omega_E^2 r'^2}{c^2}\right) (cdt')^2 + 2\omega_E r'^2 d\phi' dt' + (d\sigma')^2, \quad (4)$$

where the abbreviated expression $(d\sigma')^2 = (dr')^2 + (r'd\phi')^2 + (dz')^2$ for the square of the coordinate distance has been used.

The time transformation $t = t'$ in Eqs. (3) is deceptively simple. It means that in the rotating frame the time variable t' is really determined in the underlying inertial frame. It is an example of coordinate time. A similar concept is used in the GPS.

Now consider a process in which observers in the rotating frame attempt to use Einstein synchronization (that is, the principle of the constancy of the speed of light) to establish a network of synchronized clocks. Light travels along a null worldline, so we may set $ds^2 = 0$ in Eq. (4). Also, it is sufficient for this discussion to keep only terms of first order in the small parameter $\omega_E r'/c$. Then

$$(cdt')^2 - \frac{2\omega_E r'^2 d\phi' (cdt')}{c} - (d\sigma')^2 = 0, \quad (5)$$

and solving for (cdt') yields

$$cdt' = d\sigma' + \frac{\omega_E r'^2 d\phi'}{c}. \quad (6)$$

The quantity $r'^2 d\phi'/2$ is just the infinitesimal area dA'_z in the rotating coordinate system swept out by a vector from the rotation axis to the light pulse, and projected onto a plane parallel to the equatorial plane. Thus, the total time required for light to traverse some path is

$$\int_{\text{path}} dt' = \int_{\text{path}} \frac{d\sigma'}{c} + \frac{2\omega_E}{c^2} \int_{\text{path}} dA'_z. \quad [\text{light}] \quad (7)$$

Observers fixed on the earth, who were unaware of earth rotation, would use just $\int d\sigma'/c$ for synchronizing their clock network. Observers at rest in the underlying inertial frame would say that this leads to significant path-dependent inconsistencies, which are proportional to the projected area encompassed by the path. Consider, for example, a synchronization process that follows earth's equator in the eastwards direction. For earth, $2\omega_E/c^2 = 1.6227 \times 10^{-21} \text{ s m}^{-2}$

and the equatorial radius is $a_1 = 6,378,137$ m, so the area is $\pi a_1^2 = 1.27802 \times 10^{14}$ m². Thus, the last term in Eq. (7) is

$$\frac{2\omega_E}{c^2} \int_{\text{path}} dA'_z = 207.4 \text{ ns.} \quad (8)$$

From the underlying inertial frame, this can be regarded as the additional travel time required by light to catch up to the moving reference point. Simple-minded use of Einstein synchronization in the rotating frame gives only $\int d\sigma'/c$, and thus leads to a significant error. Traversing the equator once eastward, the last clock in the synchronization path would lag the first clock by 207.4 ns. Traversing the equator once westward, the last clock in the synchronization path would lead the first clock by 207.4 ns.

In an inertial frame a portable clock can be used to disseminate time. The clock must be moved so slowly that changes in the moving clock's rate due to time dilation, relative to a reference clock at rest on earth's surface, are extremely small. On the other hand, observers in a rotating frame who attempt this, find that the proper time elapsed on the portable clock is affected by earth's rotation rate. Factoring Eq. (4), the proper time increment $d\tau$ on the moving clock is given by

$$(d\tau)^2 = (ds/c)^2 = dt'^2 \left[1 - \left(\frac{\omega_E r'}{c} \right)^2 - \frac{2\omega_E r'^2 d\phi'}{c^2 dt'} - \left(\frac{d\sigma'}{cdt'} \right)^2 \right]. \quad (9)$$

For a slowly moving clock, $(d\sigma'/cdt')^2 \ll 1$, so the last term in brackets in Eq. (9) can be neglected. Also, keeping only first order terms in the small quantity $\omega_E r'/c$ yields

$$d\tau = dt' - \frac{\omega_E r'^2 d\phi'}{c^2} \quad (10)$$

which leads to

$$\int_{\text{path}} dt' = \int_{\text{path}} d\tau + \frac{2\omega_E}{c^2} \int_{\text{path}} dA'_z. \quad [\text{portable clock}] \quad (11)$$

This should be compared with Eq. (7). Path-dependent discrepancies in the rotating frame are thus inescapable whether one uses light or portable clocks to disseminate time, while synchronization in the underlying inertial frame using either process is self-consistent.

Eqs. (7) and (11) can be reinterpreted as a means of realizing coordinate time $t' = t$ in the rotating frame, if after performing a synchronization process appropriate corrections of the form $+2\omega_E \int_{\text{path}} dA'_z/c^2$ are applied. It is remarkable how many different ways this can be viewed. For example, from the inertial frame it appears that the reference clock from which the synchronization process starts is moving, requiring light to traverse a different path than it appears to traverse in the rotating frame. The Sagnac effect can be regarded as arising from the relativity of simultaneity in a Lorentz transformation to a sequence of local inertial frames co-moving with points on the rotating earth. It can also

be regarded as the difference between proper times of a slowly moving portable clock and a Master reference clock fixed on earth's surface.

This was recognized in the early 1980s by the Consultative Committee for the Definition of the Second and the International Radio Consultative Committee who formally adopted procedures incorporating such corrections for the comparison of time standards located far apart on earth's surface. For the GPS it means that synchronization of the entire system of ground-based and orbiting atomic clocks is performed in the local inertial frame, or ECI coordinate system [6].

GPS can be used to compare times on two earth-fixed clocks when a single satellite is in view from both locations. This is the "common-view" method of comparison of Primary standards, whose locations on earth's surface are usually known very accurately in advance from ground-based surveys. Signals from a single GPS satellite in common view of receivers at the two locations provide enough information to determine the time difference between the two local clocks. The Sagnac effect is very important in making such comparisons, as it can amount to hundreds of nanoseconds, depending on the geometry. In 1984 GPS satellites 3, 4, 6, and 8 were used in simultaneous common view between three pairs of earth timing centers, to accomplish closure in performing an around-the-world Sagnac experiment. The centers were the National Bureau of Standards (NBS) in Boulder, CO, Physikalisch-Technische Bundesanstalt (PTB) in Braunschweig, West Germany, and Tokyo Astronomical Observatory (TAO). The size of the Sagnac correction varied from 240 to 350 ns. Enough data were collected to perform 90 independent circumnavigations. The actual mean value of the residual obtained after adding the three pairs of time differences was 5 ns, which was less than 2 percent of the magnitude of the calculated total Sagnac effect [4].

3 GPS Coordinate Time and TAI

In the GPS, the time variable $t' = t$ becomes a coordinate time in the rotating frame of the earth, which is realized by applying appropriate corrections while performing synchronization processes. Synchronization is thus performed in the underlying inertial frame in which self-consistency can be achieved.

With this understanding, I next need to describe the gravitational fields near the earth due to the earth's mass itself. Assume for the moment that earth's mass distribution is static, and that there exists a locally inertial, non-rotating, freely falling coordinate system with origin at the earth's center of mass, and write an approximate solution of Einstein's field equations in isotropic coordinates:

$$-ds^2 = -\left(1 + \frac{2V}{c^2}\right)(cdt)^2 + \left(1 - \frac{2V}{c^2}\right)(dr^2 + r^2d\theta^2 + r^2\sin^2\theta d\phi^2). \quad (12)$$

where $\{r, \theta, \phi\}$ are spherical polar coordinates and where V is the Newtonian gravitational potential of the earth, given approximately by:

$$V = -\frac{GM_E}{r} \left[1 - J_2 \left(\frac{a_1}{r}\right)^2 P_2(\cos\theta)\right]. \quad (13)$$

In Eq. (13), $GM_E = 3.986004418 \times 10^{14} \text{ m}^3 \text{ s}^{-2}$ is the product of earth's mass times the Newtonian gravitational constant, $J_2 = 1.0826300 \times 10^{-3}$ is earth's quadrupole moment coefficient, and $a_1 = 6.3781370 \times 10^6$ is earth's equatorial radius¹. The angle θ is the polar angle measured downward from the axis of rotational symmetry; P_2 is the Legendre polynomial of degree 2. In using Eq. (12), it is an adequate approximation to retain only terms of first order in the small quantity V/c^2 . Higher multipole moment contributions to Eq. (13) have very a small effect for relativity in GPS.

One additional expression for the invariant interval is needed: the transformation of Eq. (12) to a rotating, ECEF coordinate system by means of transformations equivalent to Eqs. (3). The transformations for spherical polar coordinates are:

$$t = t', \quad r = r', \quad \theta = \theta', \quad \phi = \phi' + \omega_E t'. \quad (14)$$

Upon performing the transformations, and retaining only terms of order $1/c^2$, the scalar interval becomes:

$$\begin{aligned} -ds^2 = & -\left[1 + \frac{2V}{c^2} - \left(\frac{\omega_E r' \sin\theta'}{c}\right)^2\right](cdt')^2 + 2\omega_E r'^2 \sin^2\theta' d\phi' dt' \\ & + \left(1 - \frac{2V}{c^2}\right)(dr'^2 + r'^2 d\theta'^2 + r'^2 \sin^2\theta' d\phi'^2). \end{aligned} \quad (15)$$

¹WGS-84(G873) values of these constants are used in this article.

To the order of the calculation, this result is a simple superposition of the metric, Eq. (12), with the corrections due to rotation expressed in Eq. (4). The metric tensor coefficient g'_{00} in the rotating frame is

$$g'_{00} = - \left[1 + \frac{2V}{c^2} - \left(\frac{\omega_E r' \sin \theta'}{c} \right)^2 \right] \equiv - \left(1 + \frac{2\Phi}{c^2} \right), \quad (16)$$

where Φ is the effective gravitational potential in the rotating frame, which includes the static gravitational potential of the earth, and a centripetal potential term.

The Earth's geoid. In Eqs. (12) and (15), the rate of coordinate time is determined by atomic clocks at rest at infinity. The rate of GPS coordinate time, however, is closely related to International Atomic Time (TAI), which is a time scale computed by the International Bureau of Weights and Measures (BIPM) in Paris on the basis of inputs from hundreds of primary time standards, hydrogen masers, and other clocks from all over the world. In producing this time scale, corrections are applied to reduce the elapsed proper times on the contributing clocks to earth's geoid, a surface of constant effective gravitational equipotential at mean sea level in the ECEF.

Universal Coordinated Time (UTC) is another time scale, which differs from TAI by a whole number of leap seconds. These leap seconds are inserted every so often into UTC so that UTC continues to correspond to time determined by earth's rotation. Time standards organizations that contribute to TAI and UTC generally maintain their own time scales. For example, the time scale of the U.S. Naval Observatory, based on an ensemble of Hydrogen masers and Cs clocks, is denoted UTC(USNO). GPS time is steered so that, apart from the leap second differences, it stays within 100 ns UTC(USNO). Usually, this steering is so successful that the difference between GPS time and UTC(USNO) is less than about 40 ns. GPS equipment cannot tolerate leap seconds, as such sudden jumps in time would cause receivers to lose their lock on transmitted signals, and other undesirable transients would occur.

To account for the fact that reference clocks for the GPS are not at infinity, I shall consider the rates of atomic clocks at rest on the earth's geoid. These clocks move because of the earth's spin; also, they are at varying distances from the earth's center of mass since the earth is slightly oblate. In order to proceed one needs a model expression for the shape of this surface, and a value for the effective gravitational potential on this surface in the rotating frame.

For this calculation, I use Eq. (15) in the ECEF. For a clock at rest on earth, Eq. (15) reduces to

$$-ds^2 = - \left(1 + \frac{2V}{c^2} - \frac{\omega_E^2 r'^2 \sin^2 \theta'}{c^2} \right) (cdt')^2, \quad (17)$$

with the potential V given by Eq. (13). This equation determines the radius r' of the model geoid as a function of polar angle θ' . The numerical value of Φ_0

can be determined at the equator where $\theta' = \pi/2$ and $r' = a_1$. This gives

$$\begin{aligned}\frac{\Phi_0}{c^2} &= -\frac{GM_E}{a_1 c^2} - \frac{GM_E J_2}{2a_1 c^2} - \frac{\omega_E^2 a_1^2}{2c^2} \\ &= -6.95348 \times 10^{-10} - 3.764 \times 10^{-13} - 1.203 \times 10^{-12} \\ &= -6.96927 \times 10^{-10}.\end{aligned}\quad (18)$$

There are thus three distinct contributions to this effective potential: a simple $1/r$ contribution due to the earth's mass; a more complicated contribution from the quadrupole potential, and a centripetal term due to the earth's rotation. The main contribution to the gravitational potential arises from the mass of the earth; the centripetal potential correction is about 500 times smaller, and the quadrupole correction is about 2000 times smaller. These contributions have been divided by c^2 in the above equation since the time increment on an atomic clock at rest on the geoid can be easily expressed thereby. In recent resolutions of the International Astronomical Union [1], a "Terrestrial Time" scale (TT) has been defined by adopting the value $\Phi_0/c^2 = 6.969290134 \times 10^{-10}$. Eq. (18) agrees with this definition to within the accuracy needed for the GPS.

From Eq. (15), for clocks on the geoid,

$$d\tau = ds/c = dt' \left(1 + \frac{\Phi_0}{c^2} \right). \quad (19)$$

Clocks at rest on the rotating geoid run slow compared to clocks at rest at infinity by about seven parts in 10^{10} . Note that these effects sum to about 10,000 times larger than the fractional frequency stability of a high-performance Cesium clock. The shape of the geoid in this model can be obtained by setting $\Phi = \Phi_0$ and solving Eq. (16) for r' in terms of θ' . The first few terms in a power series in the variable $x' = \sin \theta'$ can be expressed as

$$r' = (6356742.025 + 21353.642 x'^2 + 39.832 x'^4 + 0.798 x'^6 + 0.003 x'^8) \text{ m.} \quad (20)$$

This treatment of the gravitational field of the oblate earth is limited by the simple model of the gravitational field. Actually, what I have done is estimate the shape of the so-called "reference ellipsoid", from which the actual geoid is conventionally measured.

Better models can be found in the literature of geophysics [18, 9, 15]. The next term in the multipole expansion of the earth's gravity field is about a thousand times smaller than the contribution from J_2 ; although the actual shape of the geoid can differ from Eq. (20) by as much as 100 meters, the effects of such terms on timing in the GPS are small. Incorporating up to 20 higher zonal harmonics in the calculation affects the value of Φ_0 only in the sixth significant figure.

Observers at rest on the geoid define the unit of time in terms of the proper rate of atomic clocks. In Eq. (19), Φ_0 is a constant. On the left side of Eq. (19), $d\tau$ is the increment of proper time elapsed on a standard clock at rest, in terms of the elapsed coordinate time dt . Thus, the very useful result has emerged,

that ideal clocks at rest on the geoid of the rotating earth all beat at the same rate. This is reasonable since the earth's surface is a gravitational equipotential surface in the rotating frame. (It is true for the actual geoid whereas I have constructed a model.) Considering clocks at two different latitudes, the one further north will be closer to the earth's center because of the flattening – it will therefore be more redshifted. However, it is also closer to the axis of rotation, and going more slowly, so it suffers less second-order Doppler shift. The earth's oblateness gives rise to an important quadrupole correction. This combination of effects cancels exactly on the reference surface.

Since all clocks at rest on the geoid beat at the same rate, it is advantageous to exploit this fact to redefine the rate of coordinate time. In Eq. (12) the rate of coordinate time is defined by standard clocks at rest at infinity. I want instead to define the rate of coordinate time by standard clocks at rest on the surface of the earth. Therefore, I shall define a new coordinate time t'' by means of a constant rate change:

$$t'' = (1 + \Phi_0/c^2)t' = (1 + \Phi_0/c^2)t. \quad (21)$$

The correction is about seven parts in 10^{10} (see Eq. (18)).

When this time scale change is made, the metric of Eq. (15) in the earth-fixed rotating frame becomes

$$\begin{aligned} -ds^2 = & - \left(1 + \frac{2(\Phi - \Phi_0)}{c^2} \right) (cdt'')^2 + 2\omega_E r'^2 \sin^2 \theta' d\phi' dt'' \\ & + \left(1 - \frac{2V}{c^2} \right) (dr'^2 + r'^2 d\theta'^2 + r'^2 \sin^2 \theta' d\phi'^2), \end{aligned} \quad (22)$$

where only terms of order c^{-2} have been retained. Whether I use dt' or dt'' in the Sagnac cross term makes no difference since the Sagnac term is very small anyway. The same time scale change in the non-rotating ECI metric, Eq. (12), gives

$$-ds^2 = - \left(1 + \frac{2(V - \Phi_0)}{c^2} \right) (cdt'')^2 + \left(1 - \frac{2V}{c^2} \right) (dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2). \quad (23)$$

Eqs. (22) and Eq. (23) imply that the proper time elapsed on clocks at rest on the geoid (where $\Phi = \Phi_0$) is identical with the coordinate time t'' . This is the correct way to express the fact that ideal clocks at rest on the geoid provide all of our standard reference clocks.

4 The Realization of Coordinate Time

We are now able to address the real problem of clock synchronization within the GPS. In the remainder of this paper I shall drop the primes on t'' and just use the symbol t , with the understanding that unit of this time is reference to UTC(USNO) on the rotating geoid, but with synchronization established in an underlying, locally inertial, reference frame. The metric Eq. (23) will henceforth be written

$$-ds^2 = -\left(1 + \frac{2(V - \Phi_0)}{c^2}\right)(cdt)^2 + \left(1 - \frac{2V}{c^2}\right)(dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2). \quad (24)$$

The difference $(V - \Phi_0)$ that appears in the first term of Eq. (24) arises because in the underlying earth-centered locally inertial (ECI) coordinate system in which Eq. (24) is expressed, the unit of time is determined by moving clocks in a spatially-dependent gravitational field.

It is obvious that Eq. (24) contains within it the well-known effects of time dilation (the apparent slowing of moving clocks) and frequency shifts due to gravitation. Due to these effects, which have an impact on the net elapsed proper time on an atomic clock, the proper time elapsing on the orbiting GPS clocks cannot be simply used to transfer time from one transmission event to another. Path-dependent effects must be accounted for.

On the other hand, according to General Relativity, the coordinate time variable t of Eq. (24) is valid in a coordinate patch large enough to cover the earth and the GPS satellite constellation. Eq. (24) is an approximate solution of the field equations near the earth, which include the gravitational fields due to earth's mass distribution. In this local coordinate patch, the coordinate time is single-valued. (It is not unique, of course, because there is still gauge freedom, but Eq. (24) represents a fairly simple and reasonable choice of gauge.) Therefore, it is natural to propose that the coordinate time variable t of Eqs. (24) and (22) be used as a basis for synchronization in the neighborhood of the earth.

To see how this works for a slowly moving atomic clock, solve Eq. (24) for dt as follows. First factor out $(cdt)^2$ from all terms on the right-hand side:

$$-ds^2 = -\left[1 + \frac{2(V - \Phi_0)}{c^2} - \left(1 - \frac{2V}{c^2}\right) \frac{dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2}{(cdt)^2}\right](cdt)^2. \quad (25)$$

I simplify by writing the velocity in the ECI coordinate system as

$$v^2 = \frac{dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2}{dt^2}. \quad (26)$$

Only terms of order c^{-2} need be kept, so the potential term modifying the velocity term can be dropped. Then, upon taking a square root, the proper time increment on the moving clock is approximately

$$d\tau = ds/c = \left[1 + \frac{(V - \Phi_0)}{c^2} - \frac{v^2}{2c^2}\right] dt. \quad (27)$$

Finally, solving for the increment of coordinate time and integrating along the path of the atomic clock,

$$\int_{\text{path}} dt = \int_{\text{path}} d\tau \left[1 - \frac{(V - \Phi_0)}{c^2} - \frac{v^2}{2c^2} \right]. \quad (28)$$

The relativistic effect on the clock, given in Eq. (27), is thus corrected by Eq. (28).

Suppose for a moment there were no gravitational fields. Then one could picture an underlying non-rotating reference frame, a local inertial frame, unattached to the spin of the earth, but with its origin at the center of the earth. In this non-rotating frame, a fictitious set of standard clocks is introduced, available anywhere, all of them being synchronized by the Einstein synchronization procedure, and running at agreed upon rates such that synchronization is maintained. These clocks read the coordinate time t . Next, one introduces the rotating earth with a set of standard clocks distributed around upon it, possibly roving around. One applies to each of the standard clocks a set of corrections based on the known positions and motions of the clocks, given by Eq. (28). This generates a “coordinate clock time” in the earth-fixed, rotating system. This time is such that at each instant the coordinate clock agrees with a fictitious atomic clock at rest in the local inertial frame, whose position coincides with the earth-based standard clock at that instant. Thus, coordinate time is equivalent to time that would be measured by standard clocks at rest in the local inertial frame [7].

When the gravitational field due to the earth is considered, the picture is only a little more complicated. There still exists a coordinate time that can be found by computing a correction for gravitational redshift, given by the first correction term in Eq. (28).

5 Relativistic Effects on Satellite Clocks

For atomic clocks in satellites, it is most convenient to consider the motions as they would be observed in the local ECI frame. Then the Sagnac effect becomes irrelevant. (The Sagnac effect on moving ground-based receivers must still be considered.) Gravitational frequency shifts and second-order Doppler shifts must be taken into account together. In this section I shall discuss in detail these two relativistic effects, using the expression for the elapsed coordinate time, Eq. (28). The term Φ_0 in Eq. (28) includes the scale correction needed in order to use clocks at rest on the earth's surface as references. The quadrupole contributes to Φ_0 in the term $-GM_E J_2/2a_1$ in Eq. (28); there it contributes a fractional rate correction of -3.76×10^{-13} . This effect must be accounted for in the GPS. Also, V is the earth's gravitational potential at the satellite's position. Fortunately, the earth's quadrupole potential falls off very rapidly with distance, and up until very recently its effect on satellite vehicle (SV) clock frequency has been neglected. This will be discussed in a later section; for the present I only note that the effect of earth's quadrupole potential on SV clocks is only about one part in 10^{14} , and I neglect it for the moment.

Satellite orbits. Let us assume that the satellites move along Keplerian orbits. This is a good approximation for GPS satellites, but poor if the satellites are at low altitude. This assumption yields relations with which to simplify Eq. (28). Since the quadrupole (and higher multipole) parts of the earth's potential are neglected, in Eq. (28) the potential is $V = -GM_E/r$. Then the expressions can be evaluated using what is known about the Newtonian orbital mechanics of the satellites. Denote the satellite's orbit semimajor axis by a and eccentricity by e . Then the solution of the orbital equations is as follows [13]: The distance r from the center of the earth to the satellite in ECI coordinates is

$$r = a(1 - e^2)/(1 + e \cos f). \quad (29)$$

The angle f , called the true anomaly, is measured from perigee along the orbit to the satellite's instantaneous position. The true anomaly can be calculated in terms of another quantity E called the eccentric anomaly, according to the relationships

$$\begin{aligned} \cos f &= \frac{\cos E - e}{1 - e \cos E}, \\ \sin f &= \sqrt{1 - e^2} \frac{\sin E}{1 - e \cos E}. \end{aligned} \quad (30)$$

Then, another way to write the radial distance r is

$$r = a(1 - e \cos E). \quad (31)$$

To find the eccentric anomaly E , one must solve the transcendental equation

$$E - e \sin E = \sqrt{\frac{GM_E}{a^3}}(t - t_p), \quad (32)$$

where t_p is the coordinate time of perigee passage.

In Newtonian mechanics, the gravitational field is a conservative field and total energy is conserved. Using the above equations for the Keplerian orbit, one can show that the total energy per unit mass of the satellite is

$$\frac{1}{2}v^2 - \frac{GM_E}{r} = -\frac{GM_E}{2a}. \quad (33)$$

If I use Eq. (33) for v^2 in Eq. (28), then I get the following expression for the elapsed coordinate time on the satellite clock:

$$\Delta t = \int_{\text{path}} d\tau \left[1 + \frac{3GM_E}{2ac^2} + \frac{\Phi_0}{c^2} - \frac{2GM_E}{c^2} \left(\frac{1}{a} - \frac{1}{r} \right) \right]. \quad (34)$$

The first two constant rate correction terms in Eq. (34) have the values:

$$\frac{3GM_E}{2ac^2} + \frac{\Phi_0}{c^2} = +2.5046 \times 10^{-10} - 6.9693 \times 10^{-10} = -4.4647 \times 10^{-10}. \quad (35)$$

The negative sign in this result means that the standard clock in orbit is beating too fast, primarily because its frequency is gravitationally blueshifted. In order for the satellite clock to appear to an observer on the geoid to beat at the chosen frequency of 10.23 MHz, the satellite clocks are adjusted lower in frequency so that the proper frequency is:

$$[1 - 4.4647 \times 10^{-10}] \times 10.23 \text{ MHz} = 10.2299999543 \text{ MHz}. \quad (36)$$

This adjustment is accomplished on the ground before the clock is placed in orbit.

Figure 2 shows the net fractional frequency offset of an atomic clock in a circular orbit, which is essentially the left side of Eq. (35) plotted as a function of orbit radius a , with a change of sign. Five sources of relativistic effects contribute in Figure 2. The effects are emphasized for several different orbit radii of particular interest. For a low earth orbiter such as the Space Shuttle, the velocity is so great that slowing due to time dilation is the dominant effect, while for a GPS satellite clock, the gravitational blueshift is greater. The effects cancel at $a \approx 9545$ km. The Global Navigation Satellite System Galileo, which is currently being designed under the auspices of the European Space Agency, will have orbital radii of approximately 30,000 km.

There is an interesting story about this frequency offset. At the time of launch of the NTS-2 satellite (23 June 1977), which contained the first Cesium atomic clock to be placed in orbit, it was recognized that orbiting clocks would require a relativistic correction, but there was uncertainty as to its magnitude as well as its sign. Indeed, there were some who doubted that relativistic effects were truths that would need to be incorporated [5]! A frequency synthesizer was built into the satellite clock system so that after launch, if in fact the rate of the clock in its final orbit was that predicted by general relativity, then the synthesizer could be turned on, bringing the clock to the coordinate rate

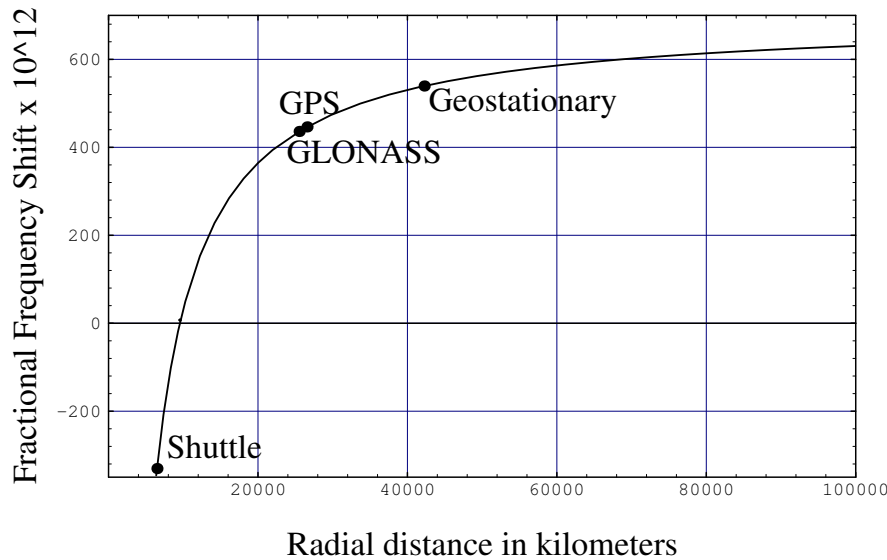


Figure 2: Net fractional frequency shift of a clock in a circular orbit.

necessary for operation. After the Cesium clock was turned on in NTS-2, it was operated for about 20 days to measure its clock rate before turning on the synthesizer [11]. The frequency measured during that interval was +442.5 parts in 10^{12} compared to clocks on the ground, while general relativity predicted +446.5 parts in 10^{12} . The difference was well within the accuracy capabilities of the orbiting clock. This then gave about a 1% verification of the combined second-order Doppler and gravitational frequency shift effects for a clock at 4.2 earth radii.

Additional small frequency offsets can arise from clock drift, environmental changes, and other unavoidable effects such as the inability to launch the satellite into an orbit with precisely the desired semimajor axis. The navigation message provides satellite clock frequency corrections for users so that in effect, the clock frequencies remain as close as possible to the frequency of the U.S. Naval Observatory's reference clock ensemble. Because of such effects, it would now be difficult to use GPS clocks to measure relativistic frequency shifts.

When GPS satellites were first deployed, the specified factory frequency offset was slightly in error because the important contribution from earth's centripetal potential (see Eq. (18)) had been inadvertently omitted at one stage of the evaluation. Although GPS managers were made aware of this error in the early 1980s, eight years passed before system specifications were changed to reflect the correct calculation [2]. As understanding of the numerous sources of error in the GPS slowly improved, it eventually made sense to incorporate the correct relativistic calculation.

The eccentricity correction. The last term in Eq. (34) may be integrated exactly by using the following expression for the rate of change of eccentric anomaly with time, which follows by differentiating Eq. (32):

$$\frac{dE}{dt} = \frac{\sqrt{GM_E/a^3}}{1 - e \cos E}. \quad (37)$$

Also, since a relativistic correction is being computed, $ds/c \simeq dt$, so

$$\begin{aligned} \int \left[\frac{2GM_E}{c^2} \left(\frac{1}{r} - \frac{1}{a} \right) \right] \frac{ds}{c} &\simeq \frac{2GM_E}{c^2} \int \left(\frac{1}{r} - \frac{1}{a} \right) dt \\ &= \frac{2GM_E}{ac^2} \int dt \left(\frac{e \cos E}{1 - e \cos E} \right) \\ &= \frac{2\sqrt{GM_E a}}{c^2} e (\sin E - \sin E_0) \\ &= + \frac{2\sqrt{GM_E a}}{c^2} e \sin E + \text{constant}. \end{aligned} \quad (38)$$

The constant of integration in Eq. (38) can be dropped since this term is lumped with other clock offset effects in the Kalman filter computation of the clock correction model. The net correction for clock offset due to relativistic effects that vary in time is

$$\Delta t_r = +4.4428 \times 10^{-10} e \sqrt{a} \sin E \frac{\text{s}}{\sqrt{\text{m}}}. \quad (39)$$

This correction must be made by the receiver; it is a correction to the coordinate time as transmitted by the satellite. For a satellite of eccentricity $e = 0.01$, the maximum size of this term is about 23 ns. The correction is needed because of a combination of effects on the satellite clock due to gravitational frequency shift and second-order Doppler shift, which vary due to orbit eccentricity.

Eq. (39) can be expressed without approximation in the alternative form

$$\Delta t_r = + \frac{2\mathbf{r} \cdot \mathbf{v}}{c^2}, \quad (40)$$

where \mathbf{r} and \mathbf{v} are the position and velocity of the satellite at the instant of transmission. This may be proved using the expressions (30, 31, 32) for the Keplerian orbits of the satellites. This latter form is usually used in implementations of the receiver software.

It is not at all necessary, in a navigation satellite system, that the eccentricity correction be applied by the receiver. It appears that the clocks in the GLONASS satellite system do have this correction applied before broadcast. In fact historically, this was dictated in the GPS by the small amount of computing power available in the early GPS satellite vehicles. It would actually make more sense to incorporate this correction into the time broadcast by the satellites; then the broadcast time events would be much closer to coordinate time – that is, GPS system time. It may now be too late to reverse this decision because of

the investment that many dozens of receiver manufacturers have in their products. However, it does mean that receivers are supposed to incorporate the relativity correction; therefore, if appropriate data can be obtained in raw form from a receiver one can measure this effect. Such measurements are discussed next.

6 TOPEX/POSEIDON Relativity Experiment

Recently, a report distributed by the Aerospace Corporation [14] claimed that the correction expressed in Eqs. (38) and (39) would not be valid for a highly dynamic receiver – *e.g.*, one in a highly eccentric orbit. This is a conceptual error, emanating from an apparently official source, which would have serious consequences. The GPS modernization program involves significant redesign and remanufacturing of the Block IIF satellites, as well as a new generation of satellites that is now on the drawing boards – the Block IIR replenishment satellites. These satellites are capable of autonomous operation, that is, they can be operated independently of the ground-based control segment for up to 180 days. They are to accomplish this by having receivers on board that determine their own position and time by listening to the other satellites that are in view. If the conceptual basis for accounting for relativity in the GPS, as it has been explained above, were invalid, the costs of opening up these satellites and reprogramming them would be astronomical.

There has been therefore considerable controversy about this issue. As a consequence, it was proposed by William Feess of the Aerospace Corporation that a measurement of this effect be made using the receiver on board the TOPEX satellite. The TOPEX satellite carries an advanced, six-channel GPS receiver. With six data channels available, five of the channels can be used to determine the bias on the local oscillator of the TOPEX receiver with some redundancy, and data from the sixth channel can be used to measure the eccentricity effect on the sixth SV clock. Here I present some preliminary results of these measurements, which are to my knowledge the first explicit measurements of the periodic part of the combined relativistic effects of time dilation and gravitational frequency shift on an orbiting receiver.

A brief description of the pseudorange measurement made by a receiver is needed here before explaining the TOPEX data. Many receivers work by generating a replica of the coded signal emanating from the transmitter. This replica, which is driven through a feedback shift register at a rate matching the Doppler-shifted incoming signal, is correlated with the incoming signal. The transmitted coordinate time can be identified in terms of a particular phase reversal at a particular point within the code train of the signal. When the correlator in the receiver is locked onto the incoming signal, the time delay between the transmission event and the arrival time, as measured on the local clock, can be measured at any chosen instant.

Let the time as transmitted from the j th satellite be denoted by t'_j . After correcting for the eccentricity effect, the GPS time of transmission would be $t'_j + (\Delta t_r)_j$. Because of SA (which was in effect for the data that were chosen), frequency offsets and frequency drifts, the satellite clock may have an additional error b_j so that the true GPS transmission time is $t_j = t'_j + (\Delta t_r)_j - b_j$.

Now the local clock, which is usually a free-running oscillator subject to various noise and drift processes, can be in error by a large amount. Let the measured reception time be t'_R and the true GPS time of reception be $t_R = t'_R - b_R$. The possible existence of this local clock bias is the reason why measurements

from four satellites are needed for navigation, as from four measurements the three components of the receiver's position vector, and the local clock bias, can be determined. The raw difference between the time of reception of the time tag from the satellite, and the time of transmission, multiplied by c , is an estimate of the geometric range between satellite and receiver called the *pseudorange* [22]:

$$\rho_j = c(t'_R - t'_j) = c[(t_R + b_R) - (t_j + b_j - (\Delta t_r)_j)]. \quad (41)$$

On the other hand the true range between satellite and receiver is

$$|\mathbf{r}_R(t_R) - \mathbf{r}_j(t_j)| = c(t_R - t_j). \quad (42)$$

Combining Eqs. (41) and (42) yields the measurement equation for this experiment:

$$|\mathbf{r}_R(t_R) - \mathbf{r}_j(t_j)| - \rho_j + cb_R - cb_j + c(\Delta t_r)_j = 0. \quad (43)$$

The purpose of the TOPEX satellite is to measure the height of the sea. This satellite has a six-channel receiver on board with a very good quartz oscillator to provide the time reference. A radar altimeter measures the distance of the satellite from the surface of the sea, but such measurements play no role in the present experiment. The TOPEX satellite has orbit radius 7,714 km, an orbital period of about 6745 seconds, and an orbital inclination of 66.06° to earth's equatorial plane. Except for perturbations due to earth's quadrupole moment, the orbit is very nearly circular, with eccentricity being only 0.000057. The TOPEX satellite is almost ideal for analysis of this relativity effect. The trajectories of the TOPEX and GPS satellites were determined independently of the on-board clocks, by means of Doppler tracking from ≈ 100 stations maintained by the Jet Propulsion Laboratory (JPL).

The receiver is a dual frequency C/A- and P-code receiver from which both code data and carrier phase data were obtained. The dual-frequency measurements enabled us to correct the propagation delay times for electron content in the ionosphere. Close cooperation was given by JPL and by William Feess in providing the dual-frequency measurements, which are ordinarily denied to civilian users, and in removing the effect of SA at time points separated by 300 seconds during the course of the experiment.

The following data were provided through the courtesy of Yoaz Bar-Sever of JPL for October 22–23, 1995:

- ECI center-of-mass position and velocity vectors for 25 satellites, in the J2000 Coordinate system with times in UTC. Data rate is every 15 minutes; accuracy quoted is 10 cm radial, 30 cm horizontal.
- ECI position and velocity vectors for the TOPEX antenna phase center. Data rate is every minute in UTC; accuracy quoted is 3 cm radial and 10 cm horizontal.
- GPS satellite clock data for 25 satellites based on ground system observations. Data rate is every 5 minutes, in GPS time; accuracy ranges between 5 and 10 cm.

- TOPEX dual frequency GPS receiver measurements of pseudorange and carrier phase for 25 satellites, a maximum of six at any one time. The data rate is every 10 seconds, in GPS time.

During this part of 1995, GPS time was ahead of UTC by 10 seconds. GPS cannot tolerate leap seconds so whenever a leap second is inserted in UTC, UTC falls farther behind GPS time. This required high-order interpolation on the orbit files to obtain positions and velocities at times corresponding to times given, every 300 seconds, in the GPS clock data files. When this was done independently by William Feess and myself we agreed typically to within a millimeter in satellite positions.

The L1 and L2 carrier phase data was first corrected for ionospheric delay. Then the corrected carrier phase data was used to smooth the pseudorange data by weighted averaging. SA was compensated in the clock data by courtesy of William Feess. Basically, the effect of SA is contained in both the clock data and in the pseudorange data and can be eliminated by appropriate subtraction. Corrections for the offset of the GPS SV antenna phase centers from the SV centers of mass were also incorporated.

The determination of the TOPEX clock bias is obtained by rearranging Eq. (43):

$$|\mathbf{r}_R(t_R) - \mathbf{r}_j(t_j)| - \rho_j - cb_j + c\Delta t_r = -cb_R. \quad (44)$$

Generally, at each time point during the experiment, observations were obtained from six (sometimes five) satellites. The geometric range, the first term in Eq. (44), was determined by JPL from independent Doppler tracking of both the GPS constellation and the TOPEX satellite. The pseudorange was directly measured by the receiver, and clock models provided the determination of the clock biases cb_j in the satellites. The relativity correction for each satellite can be calculated directly from the given GPS satellite orbits. Because the receiver is a six-channel receiver, there is sufficient redundancy in the measurements to obtain good estimates of the TOPEX clock bias and the rms error in this bias due to measurement noise. The resulting clock bias is plotted in Figure 3.

The rms deviation from the mean of the TOPEX clock biases is plotted in Figure 4 as a function of time. The average rms error is 29 cm, corresponding to about one ns of propagation delay. Much of this variation can be attributed to multipath effects.

Figure 3 shows an overall frequency drift, accompanied by frequency adjustments and a large periodic variation with period equal to the orbital period. Figure 3 gives our best estimate of the TOPEX clock bias. This may now be used to measure the eccentricity effects by rearranging Eq. (43):

$$|\mathbf{r}_R(t_R) - \mathbf{r}_j(t_j)| - \rho_j - cb_j + cb_R = -c\Delta t_r. \quad (45)$$

Strictly speaking, in finding the eccentricity effect this way for a particular satellite, one should not include data from that satellite in the determination of the clock bias. One can show, however, that the penalty for this is simply to increase the rms error by a factor of 6/5, to 35 cm. Figure 5 shows the measured

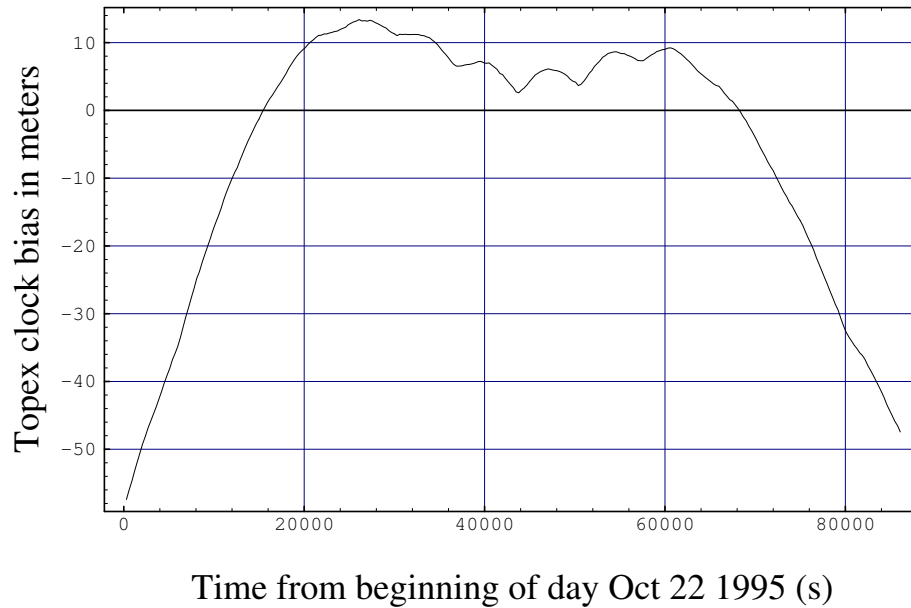


Figure 3: *TOPEX clock bias in meters determined from 1,571 observations.*

eccentricity effect for SV nr. 13, which has the largest eccentricity of the satellites that were tracked, $e = 0.01486$. The solid curve in Figure 5 is the theoretically predicted effect, from Eq. (39). While the agreement is fairly good, one can see some evidence of systematic bias during particular passes, where the rms error (plotted as vertical lines on the measured dots) is significantly smaller than the discrepancies between theory and experiment. For this particular satellite, the rms deviation between theory and experiment is 22 cm, which is about 2.2% of the maximum magnitude of the effect, 10.2 m.

Similar plots were obtained for 25 GPS satellites that were tracked during this experiment. Rather than show them one by one, it is interesting to plot them on the same graph by dividing the calculated and measured values by eccentricity e , while translating the time origin so that in each case time is measured from the instant of perigee passage. We plot the effects, not the corrections. In this way, Figure 6 combines the eccentricity effects for the five satellites with the largest eccentricities. These are SV's nr. 13, 21, 27, 23, and 26. In Figure 6 the systematic deviations between theory and experiment tend to occur for one satellite during a pass; this “pass bias” might be removable if we understood better what the cause of it is. As it stands, the agreement between theory and experiment is within about 2.5%.

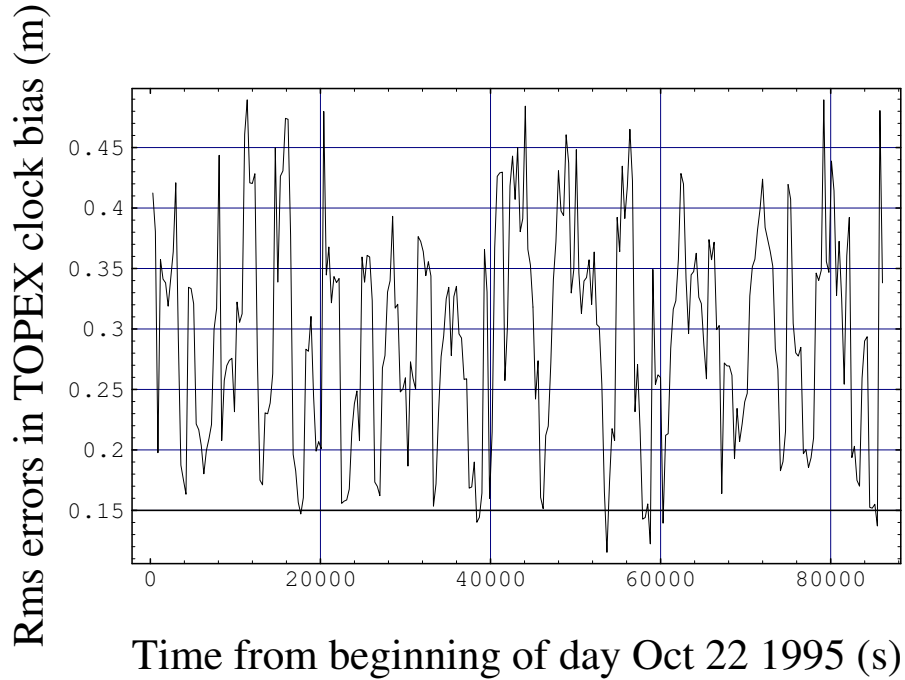


Figure 4: *Rms deviation from mean of TOPEX clock bias determinations.*

7 Doppler Effect

Since orbiting clocks have had their rate adjusted so that they beat coordinate time, and since responsibility for correcting for the periodic relativistic effect due to eccentricity has been delegated to receivers, one must take extreme care in discussing the Doppler effect for signals transmitted from satellites. Even though second-order Doppler effects have been accounted for, for earth-fixed users there will still be a first-order (longitudinal) Doppler shift, which has to be dealt with by receivers. As is well known, in a static gravitational field coordinate frequency is conserved during propagation of an electromagnetic signal along a null geodesic. If one takes into account only the monopole and quadrupole contributions to earth's gravitational field, then the field is static and one can exploit this fact to discuss the Doppler effect.

Consider the transmission of signals from rate-adjusted transmitters orbiting on GPS satellites. Let the gravitational potential and velocity of the satellite be $V(\mathbf{r}_j) \equiv V_j$, and \mathbf{v}_j , respectively. Let the frequency of the satellite transmission, before the rate adjustment is done, be $f_0 = 10.23$ MHz. After taking into account the rate adjustment discussed previously, it is straightforward to show that for a receiver of velocity \mathbf{v}_R and gravitational potential V_R (in ECI

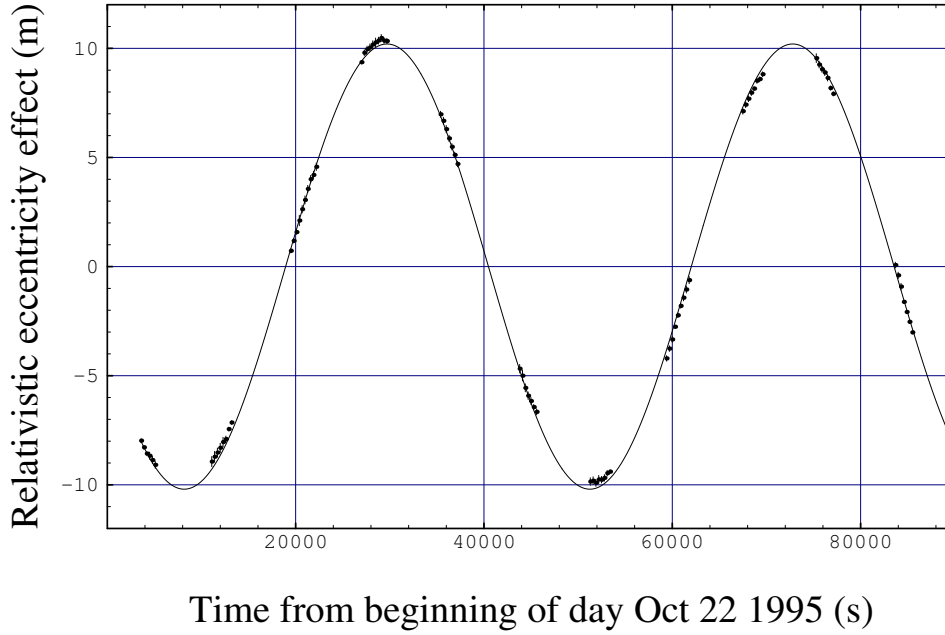


Figure 5: *Comparison of predicted and measured eccentricity effect for SV nr. 13.*

coordinates), the received frequency is

$$f_R = f_0 \left[1 + \frac{-V_R + \mathbf{v}_R^2/2 + \Phi_0 + 2GM_E/a + 2V_j}{c^2} \right] \frac{(1 - \mathbf{N} \cdot \mathbf{v}_R/c)}{(1 - \mathbf{N} \cdot \mathbf{v}_j/c)}, \quad (46)$$

where \mathbf{N} is a unit vector in the propagation direction in the local inertial frame. For a receiver fixed on the earth's rotating geoid, this reduces to

$$f_R = f_0 \left[1 + \frac{2GM_E}{c^2} \left(\frac{1}{a} - \frac{1}{r} \right) \right] \frac{(1 - \mathbf{N} \cdot \mathbf{v}_R/c)}{(1 - \mathbf{N} \cdot \mathbf{v}_j/c)}. \quad (47)$$

The correction term in square brackets gives rise to the eccentricity effect. The longitudinal Doppler shift factors are not affected by these adjustments; they will be of order 10^{-5} while the eccentricity effect is of order $e \times 10^{-10}$.

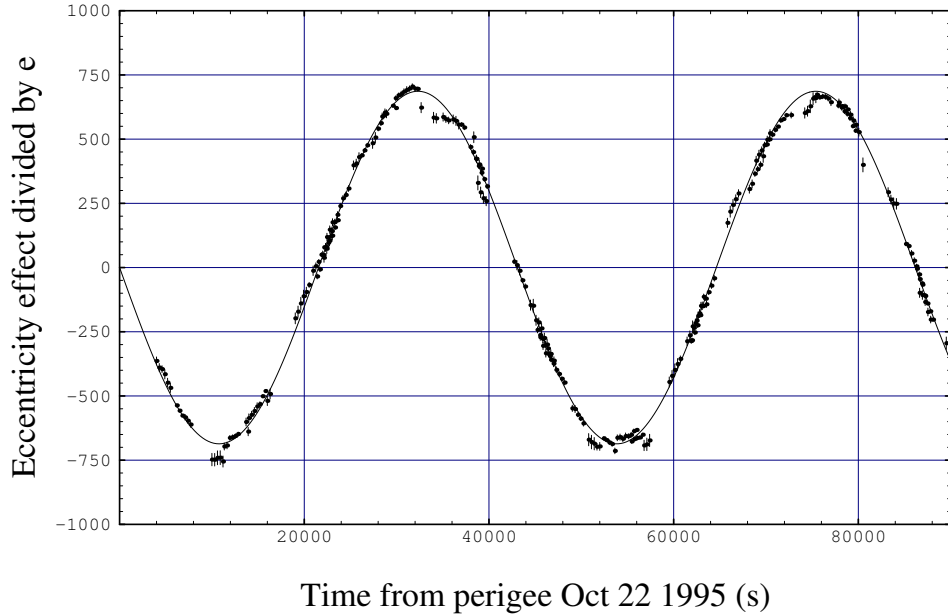


Figure 6: *Generic eccentricity effect for five satellites.*

8 Crosslink Ranging

Consider next the process of transferring coordinate time from one satellite clock to another by direct exchange of signals. This will be important when “Autonav” is implemented. The standard atomic clock in the transmitter satellite suffers a rate adjustment, and an eccentricity correction to get the coordinate time. Then a signal is sent to the second satellite which requires calculating a coordinate time of propagation possibly incorporating a relativistic time delay. There is then a further transformation of rate and another “ $e \sin E$ ” correction to get the atomic time on the receiving satellite’s clock. So that the rate adjustment does not introduce confusion into this analysis, I shall assume the rate adjustments are already accounted for and use the subscript ‘S’ to denote coordinate time measurements using rate-adjusted satellite clocks.

Then, let a signal be transmitted from satellite nr. i , at position \mathbf{r}_i and having velocity \mathbf{v}_i in ECI coordinates, at satellite clock time $T_S^{(i)}$, to satellite nr. j , at position \mathbf{r}_j and having velocity \mathbf{v}_j . The coordinate time at which this occurs, apart from a constant offset, from Eq. (38), will be

$$T^{(i)} = T_S^{(i)} + \frac{2\sqrt{GMa_i}}{c^2} e_i \sin E_i. \quad (48)$$

The coordinate time elapsed during propagation of the signal to the receiver in satellite nr. j is in first approximation l/c , where l is the distance between transmitter at the instant of transmission, and receiver at the instant of recep-

tion: $\Delta T = T^{(j)} - T^{(i)} = l/c$. The Shapiro time delay corrections to this will be discussed in the next section. Finally, the coordinate time of arrival of the signal is related to the time on the receiving satellite's adjusted clock by the inverse of Eq. (48):

$$T_S^{(j)} = T^{(j)} - \frac{2\sqrt{GMa_j}}{c^2} e_j \sin E_j. \quad (49)$$

Collecting these results, we get

$$T_S^{(j)} = T_S^{(i)} + \frac{l}{c} - \frac{2\sqrt{GMa_j}}{c^2} e_j \sin E_j + \frac{2\sqrt{GMa_i}}{c^2} e_i \sin E_i. \quad (50)$$

In Eq. (50) the distance l is the actual propagation distance, in ECI coordinates, of the signal. If this is expressed instead in terms of the distance $|\Delta \mathbf{r}| = |\mathbf{r}_j(t_i) - \mathbf{r}_i(t_i)|$ between the two satellites at the instant of transmission, then

$$l = |\Delta \mathbf{r}| + \frac{\Delta \mathbf{r} \cdot \mathbf{v}_j}{c}. \quad (51)$$

The extra term accounts for motion of the receiver through the inertial frame during signal propagation. Then Eq. (50) becomes

$$T_S^{(j)} = T_S^{(i)} + \frac{|\Delta \mathbf{r}|}{c} - \frac{2\sqrt{GMa_j}}{c^2} e_j \sin E_j + \frac{2\sqrt{GMa_i}}{c^2} e_i \sin E_i + \frac{\Delta \mathbf{r} \cdot \mathbf{v}_j}{c^2}. \quad (52)$$

This result contains all the relativistic corrections that need to be considered for direct time transfer by transmission of a time-tagged pulse from one satellite to another. The last term in Eq. (52) should *not* be confused with the correction of Eq. (40).

9 Frequency Shifts Induced by Orbit Changes

Improvements in GPS motivate attention to other small relativistic effects that have previously been too small to be explicitly considered. For SV clocks, these include frequency changes due to orbit adjustments, and effects due to earth's oblateness. For example, between July 25 and October 10, 2000, SV43 occupied a transfer orbit while it was moved from slot 5 to slot 3 in orbit plane F. I will show here that the fractional frequency change associated with a change da in the semi-major axis a (in meters) can be estimated as $9.429 \times 10^{-18} da$. In the case of SV43, this yields a prediction of -1.77×10^{-13} for the fractional frequency change of the SV43 clock which occurred 25 July 2000. This relativistic effect was recently measured very carefully [12]. Another orbit adjustment on October 10, 2000 should have resulted in another fractional frequency change of $+1.75 \times 10^{-13}$, which has not yet been measured carefully. Also, earth's oblateness causes a periodic fractional frequency shift with period of almost 6 hours and amplitude 0.695×10^{-14} . This means that quadrupole effects on SV clock frequencies are of the same general order of magnitude as the frequency breaks induced by orbit changes. Thus, some approximate expressions for the frequency effects on SV clock frequencies due to earth's oblateness are needed. These effects will be discussed with the help of Lagrange's planetary perturbation equations.

Five distinct relativistic effects, discussed in Section 5, are incorporated into the System Specification Document, ICD-GPS-200 [2]. These are:

- the effect of earth's mass on gravitational frequency shifts of atomic reference clocks fixed on the earth's surface relative to clocks at infinity;
- the effect of earth's oblate mass distribution on gravitational frequency shifts of atomic clocks fixed on earth's surface;
- second-order Doppler shifts of clocks fixed on earth's surface due to earth rotation;
- gravitational frequency shifts of clocks in GPS satellites due to earth's mass;
- and second-order Doppler shifts of clocks in GPS satellites due to their motion through an Earth-Centered Inertial (ECI) Frame.

The combination of second-order Doppler and gravitational frequency shifts given in Eq. (27) for a clock in a GPS satellite leads directly to the following expression for the fractional frequency shift of a satellite clock relative to a reference clock fixed on earth's geoid:

$$\frac{\Delta f}{f} = -\frac{1}{2} \frac{v^2}{c^2} - \frac{GM_E}{rc^2} - \frac{\Phi_0}{c^2}, \quad (53)$$

where v is the satellite speed in a local ECI reference frame, GM_E is the product of the Newtonian gravitational constant G and earth's mass M , c is the defined

speed of light, and Φ_0 is the effective gravitational potential on the earth's rotating geoid. The term Φ_0 includes contributions from both monopole and quadrupole moments of earth's mass distribution, and the effective centripetal potential in an earth-fixed reference frame such as the WGS-84(837) frame, due to earth's rotation. The value for Φ_0 is given in Eq. (18), and depends on earth's equatorial radius a_1 , earth's quadrupole moment coefficient J_2 , and earth's angular rotational speed ω_E .

If the GPS satellite orbit can be approximated by a Keplerian orbit of semi-major axis a , then at an instant when the distance of the clock from earth's center of mass is r , this leads to the following expression for the fraction frequency shift of Eq. (53):

$$\frac{\Delta f}{f} = -\frac{3GM_E}{2ac^2} - \frac{\Phi_0}{c^2} + \frac{2GM_E}{c^2} \left[\frac{1}{r} - \frac{1}{a} \right]. \quad (54)$$

Eq. (54) is derived by making use of the conservation of total energy (per unit mass) of the satellite, Eq. (33), which leads to an expression for v^2 in terms of GM_E/r and GM_E/a that can be substituted into Eq. (53). The first two terms in Eq. (54) give rise to the “factory frequency offset”, which is applied to GPS clocks before launch in order to make them beat at a rate equal to that of reference clocks on earth's surface. The last term in Eq. (54) is very small when the orbit eccentricity e is small; when integrated over time these terms give rise to the so-called “ $e \sin E$ ” effect or “eccentricity effect”. In most of the following discussion we shall assume that eccentricity is very small.

Clearly, from Eq. (54), if the semi-major axis should change by an amount δa due to an orbit adjustment, the satellite clock will experience a fractional frequency change

$$\frac{\delta f}{f} = +\frac{3GM_E \delta a}{2c^2 a^2}. \quad (55)$$

The factor 3/2 in this expression arises from the combined effect of second-order Doppler and gravitational frequency shifts. If the semi-major axis increases, the satellite will be higher in earth's gravitational potential and will be gravitationally blue-shifted more, while at the same time the satellite velocity will be reduced, reducing the size of the second-order Doppler shift (which is generally a red shift). The net effect would make a positive contribution to the fractional frequency shift.

Although it has long been known that orbit adjustments are associated with satellite clock frequency shifts, nothing has been documented and up until recently no reliable measurements of such shifts have been made. On July 25, 2000, a trajectory change was applied to SV43 to shift the satellite from slot F5 to slot F3. A drift orbit extending from July 25, 2000 to October 10, 2000 was used to accomplish this move. A “frequency break” was observed but the cause of this frequency jump was not initially understood. Recently, Marvin Epstein, Joseph Fine, and Eric Stoll [12] of ITT have evaluated the frequency shift of SV43 arising from this trajectory change. They reported that associated with the thruster firings on July 25, 2000 there was a frequency shift of the Rubidium

clock on board SV43 of amount

$$\frac{\delta f}{f} = -1.85 \times 10^{-13} \quad (\text{measured}). \quad (56)$$

Epstein *et al.* [12] suggested that the above frequency shift was relativistic in origin, and used precise ephemerides obtained from the National Imagery and Mapping Agency to estimate the frequency shift arising from second-order Doppler and gravitational potential differences. They calculated separately the second-order Doppler and gravitational frequency shifts due to the orbit change. The NIMA precise ephemerides are expressed in the WGS-84 coordinate frame, which is earth-fixed. If used without removing the underlying earth rotation, the velocity would be erroneous. They therefore transformed the NIMA precise ephemerides to an earth-centered inertial frame by accounting for a (uniform) earth rotation rate.

The semi-major axes before and after the orbit change were calculated by taking the average of the maximum and minimum radial distances. Speeds were calculated using a Keplerian orbit model. They arrived at the following numerical values for semi-major axis and velocity:

$$\begin{aligned} 07/22/00 : a &= 2.656139556 \times 10^7 \text{ m}; & v &= 3.873947951 \times 10^3 \text{ m s}^{-1}, \\ 07/30/00 : a &= 2.654267359 \times 10^7 \text{ m}; & v &= 3.875239113 \times 10^3 \text{ m s}^{-1}. \end{aligned}$$

Since the semi-major axis decreased, the frequency shift should be negative. The prediction they made for the frequency shift, which was based on Eq. (53), was then

$$\frac{\delta f}{f} = -1.734 \times 10^{-13}, \quad (57)$$

which is to be compared with the measured value, Eq. (56). This is fairly compelling evidence that the observed frequency shift is indeed a relativistic effect.

Lagrange perturbation theory. Perturbations of GPS orbits due to earth's quadrupole mass distribution are a significant fraction of the change in semi-major axis associated with the orbit change discussed above. This raises the question whether it is sufficiently accurate to use a Keplerian orbit to describe GPS satellite orbits, and estimate the semi-major axis change as though the orbit were Keplerian. In this section, we estimate the effect of earth's quadrupole moment on the orbital elements of a nominally circular orbit and thence on the change in frequency induced by an orbit change. Previously, such an effect on the SV clocks has been neglected, and indeed it does turn out to be small. However, the effect may be worth considering as GPS clock performance continues to improve.

To see how large such quadrupole effects may be, we use exact calculations for the perturbations of the Keplerian orbital elements available in the literature [13]. For the semi-major axis, if the eccentricity is very small, the

dominant contribution has a period twice the orbital period and has amplitude $3J_2a_1^2\sin^2 i_0/(2a_0) \approx 1658$ m. WGS-84(837) values for the following additional constants are used in this section: $a_1 = 6.3781370 \times 10^6$ m; $\omega_E = 7.291151467 \times 10^{-5}$ s $^{-1}$; $a_0 = 2.656175 \times 10^7$ m, where a_1 and a_0 are earth's equatorial radius and SV orbit semi-major axis, and ω_E is earth's rotational angular velocity.

The oscillation in the semi-major axis would significantly affect calculations of the semi-major axis at any particular time. This suggests that Eq. (33) needs to be reexamined in light of the periodic perturbations on the semi-major axis. Therefore, in this section we develop an approximate description of a satellite orbit of small eccentricity, taking into account earth's quadrupole moment to first order. Terms of order $J_2 \times e$ will be neglected. This problem is non-trivial because the perturbations themselves (see, for example, the equations for mean anomaly and altitude of perigee) have factors $1/e$ which blow up as the eccentricity approaches zero. This problem is a mathematical one, not a physical one. It simply means that the observable quantities – such as coordinates and velocities – need to be calculated in such a way that finite values are obtained. Orbital elements that blow up are unobservable.

Conservation of energy. The gravitational potential of a satellite at position (x, y, z) in equatorial ECI coordinates in the model under consideration here is

$$V(x, y, z) = -\frac{GM_E}{r} \left(1 - \frac{J_2 a_1^2}{r^2} \left[\frac{3z^2}{2r^2} - \frac{1}{2} \right] \right). \quad (58)$$

Since the force is conservative in this model (solar radiation pressure, thrust, *etc.* are not considered), the kinetic plus potential energy is conserved. Let ϵ be the energy per unit mass of an orbiting mass point. Then

$$\epsilon = \text{constant} = \frac{v^2}{2} + V(x, y, z) = \frac{v^2}{2} - \frac{GM_E}{r} + V'(x, y, z), \quad (59)$$

where $V'(x, y, z)$ is the perturbing potential due to the earth's quadrupole potential. It is shown in textbooks [13] that, with the help of Lagrange's planetary perturbation theory, the conservation of energy condition can be put in the form

$$\epsilon = -\frac{GM_E}{2a} + V'(x, y, z), \quad (60)$$

where a is the perturbed (osculating) semi-major axis. In other words, for the perturbed orbit,

$$\frac{v^2}{2} - \frac{GM_E}{r} = -\frac{GM_E}{2a}. \quad (61)$$

On the other hand, the net fractional frequency shift relative to a clock at rest at infinity is determined by the second-order Doppler shift (a red shift) and a gravitational redshift. The total relativistic fractional frequency shift is

$$\frac{\Delta f}{f} = -\frac{v^2}{2} - \frac{GM_E}{r} + V'(x, y, z). \quad (62)$$

The conservation of energy condition can be used to express the second-order Doppler shift in terms of the potential. Since in this paper we are interested in fractional frequency changes caused by changing the orbit, it will make no difference if the calculations use a clock at rest at infinity as a reference rather than a clock at rest on earth's surface. The reference potential cancels out to the required order of accuracy. Therefore, from perturbation theory we need expressions for the square of the velocity, for the radius r , and for the perturbing potential. We now proceed to derive these expressions. We refer to the literature [13] for the perturbed osculating elements. These are exactly known, to all orders in the eccentricity, and to first order in J_2 . We shall need only the leading terms in eccentricity e for each element.

Perturbation equations. First we recall some facts about an unperturbed Keplerian orbit, which have already been introduced (see Section 5). The eccentric anomaly E is to be calculated by solving the equation

$$E - e \sin E = M = n_0(t - t_0), \quad (63)$$

where M is the “mean anomaly” and t_0 is the time of passage past perigee, and

$$n_0 = \sqrt{GM_E/a^3}. \quad (64)$$

Then, the perturbed radial distance r and true anomaly f of the satellite are obtained from

$$r = a(1 - e \cos E), \quad (65)$$

$$\cos f = \frac{\cos E - e}{1 - e \cos E}, \quad \sin f = \sqrt{1 - e^2} \frac{\sin E}{1 - e \cos E}. \quad (66)$$

The observable x, y, z -coordinates of the satellite are then calculated from the following equations:

$$x = r(\cos \Omega \cos(f + \omega) - \cos i \sin \Omega \sin(f + \omega)), \quad (67)$$

$$y = r(\sin \Omega \cos(f + \omega) + \cos i \cos \Omega \sin(f + \omega)), \quad (68)$$

$$z = r(\sin i \sin(f + \omega)), \quad (69)$$

where Ω is the angle of the ascending line of nodes, i is the inclination, and ω is the altitude of perigee. By differentiation with respect to time, or by using the conservation of energy equation, one obtains the following expression for the square of the velocity:

$$v^2 = \frac{GM_E}{a} \frac{1 + e \cos E}{1 - e \cos E}. \quad (70)$$

In these expressions v^2 and r^{-1} are observable quantities. The combination $e \cos E$, where E is the eccentric anomaly, occurs in both of these expressions. To derive expressions for v^2 and r^{-1} in the perturbed orbits, expressions for the perturbed elements a , e , E are to be substituted into the right-hand sides of the Keplerian equations for E , r , and v^2 . Therefore, we need the combination $e \cos E$ in the limit of small eccentricity.

Perturbed eccentricity. To leading order, from the literature [13] we have for the perturbed eccentricity the following expression:

$$e = e_0 + \frac{3J_2a_1^2}{2a_0^2} \left[\left(1 - \frac{3}{2} \sin^2 i_0 \right) \cos f + \frac{1}{4} \sin^2 i_0 \cos(2\omega_0 + f) + \frac{7}{12} \sin^2 i_0 \cos(2\omega_0 + 3f) \right], \quad (71)$$

where e_0 is a constant of integration.

Perturbed eccentric anomaly. The eccentric anomaly is calculated from the equation

$$E = M + e \sin E, \quad (72)$$

with perturbed values for M and e . Expanding to first order in e gives the following expression for $\cos E$:

$$\cos E = \cos M - e \sin M \sin E, \quad (73)$$

and multiplying by e yields

$$e \cos E = e \cos M - e^2 \sin M \sin E \approx e \cos M. \quad (74)$$

We shall neglect higher order terms in e . The perturbed expression for mean anomaly M can be written as

$$M = M_0 + \Delta M/e_0, \quad (75)$$

where we indicate explicitly the terms in e_0^{-1} ; that is, the quantity M_0 contains all terms which do not blow up as $e \rightarrow 0$, and $\Delta M/e_0$ contains all the other terms. The perturbations of M are known exactly but we shall only need the leading order terms, which are

$$\Delta M/e_0 = -\frac{3J_2a_1^2}{2e_0a_0^2} \left[\left(1 - \frac{3}{2} \sin^2 i_0 \right) \sin f - \frac{1}{4} \sin^2 i_0 \sin(2\omega_0 + f) + \frac{7}{12} \sin^2 i_0 \sin(2\omega_0 + 3f) \right], \quad (76)$$

and so for very small eccentricity,

$$e \cos E = e \cos M_0 - \Delta M \sin M_0. \quad (77)$$

Then after accounting for contributions from the perturbed eccentricity and the perturbed mean anomaly, after a few lines of algebra we obtain the following for $e \cos E$:

$$e \cos E = e_0 \cos E_0 + \frac{3J_2a_1^2}{2a_0^2} \left(1 - \frac{3}{2} \sin^2 i_0 \right) + \frac{5J_2a_1^2}{4a_0^2} \sin^2 i_0 \cos 2(\omega_0 + f), \quad (78)$$

where the first term is the unperturbed part. The perturbation is a constant, plus a term with twice the orbital period.

Perturbation in semi-major axis. From the literature, the leading terms in the perturbation of the semi-major axis are

$$a = a_0 + \frac{3J_2a_1^2}{2a_0} \sin^2 i_0 \cos 2(\omega_0 + f), \quad (79)$$

where a_0 is a constant of integration. The amplitude of the periodic term is about 1658 meters.

Perturbation in radius. We are now in position to compute the perturbation in the radius. From the expression for r , after combining terms we have

$$\begin{aligned} r &= a_0(1 - e_0 \cos E_0) + \Delta a - \Delta(e \cos E) \\ &= a_0(1 - e_0 \cos E_0) - \frac{3J_2a_1^2}{2a_0} \left(1 - \frac{3}{2} \sin^2 i_0\right) \\ &\quad + \frac{J_2a_1^2}{4a_0} \sin^2 i_0 \cos 2(\omega_0 + f). \end{aligned} \quad (80)$$

The amplitude of the periodic part of the perturbation in the observable radial distance is only 276 meters.

Perturbation in the velocity squared. The above results, after substituting into Eq. (70), yield the expression

$$\begin{aligned} \frac{v^2}{2} &= \frac{GM_E}{2a_0} (1 + 2e_0 \cos E_0) + \frac{3GM_E J_2 a_1^2}{a_0^3} \left(1 - \frac{3}{2} \sin^2 i_0\right) \\ &\quad + \frac{GM_E J_2 a_1^2}{2a_0^3} \sin^2 i_0 \cos 2(\omega_0 + f). \end{aligned} \quad (81)$$

Perturbation in GM_E/r . The above expression for the perturbed r yields the following for the monopole contribution to the gravitational potential:

$$\begin{aligned} -\frac{GM_E}{r} &= -\frac{GM_E}{a_0} (1 + e_0 \cos E_0) - \frac{3GM_E J_2 a_1^2}{2a_0^3} \left(1 - \frac{3}{2} \sin^2 i_0\right) \\ &\quad + \frac{GM_E J_2 a_1^2 \sin^2 i_0}{4a_0^3} \cos 2(\omega_0 + f). \end{aligned} \quad (82)$$

Evaluation of the perturbing potential. Since the perturbing potential contains the small factor J_2 , to leading order we may substitute unperturbed values for r and z into $V'(x, y, z)$, which yields the expression

$$V'(x, y, z) = -\frac{GM_E J_2 a_1^2}{2a_0^3} \left(1 - \frac{3}{2} \sin^2 i_0\right) - \frac{3GM_E J_2 a_1^2 \sin^2 i_0}{4a_0^3} \cos 2(\omega_0 + f). \quad (83)$$

Conservation of energy. It is now very easy to check conservation of energy. Adding kinetic energy per unit mass to two contributions to the potential energy gives

$$\epsilon = \frac{v^2}{2} - \frac{GM_E}{r} + V' = -\frac{GM_E}{2a_0} - \frac{GM_E J_2 a_1^2}{2a_0^3} \left(1 - \frac{3}{2} \sin^2 i_0\right). \quad (84)$$

This verifies that the perturbation theory gives a constant energy. The extra term in the above equation, with J_2 in it, can be neglected. This is because the nominal inclination of GPS orbits is such that the factor $(1 - 3 \sin^2 i_0/2)$ is essentially zero. The near vanishing of this factor is pure coincidence in the GPS. There was no intent, in the original GPS design, that quadrupole effects would be simpler if the orbital inclination were close to 55° . However, because this term is negligible, numerical calculations of the total energy per unit mass provide a means of evaluating the quantity a_0 .

Calculation of fractional frequency shift. The fractional frequency shift calculation is very similar to the calculation of the energy, except that the second-order Doppler term contributes with a *negative* sign. The result is

$$\begin{aligned} \frac{\Delta f}{f} &= -\frac{v^2}{2c^2} - \frac{GM_E}{c^2 r} + \frac{V'}{c^2} \\ &= -\frac{3GM_E}{2a_0 c^2} - \frac{2GM_E}{c^2 a_0} e_0 \cos E_0 - \frac{7GM_E J_2 a_1^2}{2a_0^3 c^2} \left(1 - \frac{3}{2} \sin^2 i_0\right) \\ &\quad - \frac{GM_E J_2 a_1^2 \sin^2 i_0}{a_0^3 c^2} \cos 2(\omega_0 + f). \end{aligned} \quad (85)$$

The first term, when combined with the reference potential at earth's geoid, gives rise to the “factory frequency offset”. The second term gives rise to the eccentricity effect. The third term can be neglected, as pointed out above. The last term has an amplitude

$$\frac{GM_E J_2 a_1^2 \sin^2 i_0}{a_0^3 c^2} = 6.95 \times 10^{-15}, \quad (86)$$

which may be large enough to consider when calculating frequency shifts produced by orbit changes. Therefore, this contribution may have to be considered in the future in the determination of the semi-major axis, but for now we neglect it.

The result suggests the following method of computing the fractional frequency shift: Averaging the shift over one orbit, the periodic term will average down to a negligible value. The third term is negligible. So if one has a good estimate for the nominal semi-major axis parameter, the term $-3GM_E/2a_0 c^2$ gives the average fractional frequency shift. On the other hand, the average energy per unit mass is given by $\epsilon = -GM_E/2a_0$. Therefore, the precise ephemerides, specified in an ECI frame, can be used to compute the average value for ϵ ; then

the average fractional frequency shift will be

$$\frac{\Delta f}{f} = 3\epsilon/c^2. \quad (87)$$

The last periodic term in Eq. (85) is of a form similar to that which gives rise to the eccentricity correction, which is applied by GPS receivers. Considering only the last periodic term, the additional time elapsed on the orbiting clock will be given by

$$\delta t_{J_2} = \int_{\text{path}} dt \left[-\frac{GM_E J_2 a_1^2 \sin^2 i_0}{a_0^3 c^2} \cos(2\omega_0 + 2nt) \right], \quad (88)$$

where to a sufficient approximation we have replaced the quantity f in the integrand by $n = \sqrt{GM_E/a_0^3}$; n is the approximate mean motion of GPS satellites. Integrating and dropping the constant of integration (assuming as usual that such constant time offsets are lumped with other contributions) gives the periodic relativistic effect on the elapsed time of the SV clock due to earth's quadrupole moment:

$$\delta t_{J_2} = -\sqrt{\frac{GM_E}{a_0^3}} \frac{J_2 a_1^2 \sin^2 i_0}{2c^2} \sin(2\omega_0 + 2nt). \quad (89)$$

The correction that should be applied by the receiver is the *negative* of this expression,

$$\delta t_{J_2}(\text{correction}) = \sqrt{\frac{GM_E}{a_0^3}} \frac{J_2 a_1^2 \sin^2 i_0}{2c^2} \sin(2\omega_0 + 2nt). \quad (90)$$

The phase of this correction is zero when the satellite passes through earth's equatorial plane going northwards. If not accounted for, this effect on the SV clock time would give rise to a peak-to-peak periodic navigational error in position of approximately $2c \times \delta t_{J_2} = 1.43$ cm.

These effects were considered by Ashby and Spilker [9], pp. 685–686, but in that work the effect of earth's quadrupole moment on the term GM_E/r was not considered; the present calculations supercede that work.

Numerical calculations. Precise ephemerides were obtained for SV43 from the web site “ftp://sideshow.jpl.nasa.gov/pub/gipsy_products/2000/orbits” at the Jet Propulsion Laboratory. These are expressed in the J2000 ECI frame. Computer code was written to compute the average value of ϵ for one day and thence the fractional frequency shift relative to infinity before and after each orbit change. The following results were obtained:

$$\begin{aligned} 07/22/00 : a &= (2.65611575 \times 10^7 \pm 69) \text{ m}, \\ 07/30/00 : a &= (2.65423597 \times 10^3 \pm 188) \text{ m}, \\ 10/07/00 : a &= (2.65418742 \times 10^7 \pm 95) \text{ m}, \\ 10/12/00 : a &= (2.65606323 \times 10^7 \pm 58) \text{ m}. \end{aligned}$$

Therefore, the fractional frequency change produced by the orbit change of July 25 is calculated to be

$$\frac{\Delta f}{f} = -1.77 \times 10^{-13}, \quad (91)$$

which agrees with the measured value to within about 3.3%. The agreement is slightly better than that obtained in [12], perhaps because they did not consider contributions to the energy from the quadrupole moment term.

A similar calculation shows that the fractional frequency shift of SV43 on October 10, 2001 should have been

$$\frac{\Delta f}{f} = +1.75 \times 10^{-13}. \quad (92)$$

This shift has not yet been measured accurately.

On March 9, 2001, SV54's orbit was changed by firing the thruster rockets. Using the above procedures, I can calculate the fractional frequency change produced in the onboard clocks. The result is

$$\begin{aligned} 03/07/01 : a &= (2.65597188 \times 10^7 \pm 140) \text{ m}, \\ 03/11/01 : a &= (2.65359261 \times 10^7 \pm 108) \text{ m}. \end{aligned}$$

Using Eq. (55) yields the following prediction for the fractional frequency change of SV54 on March 9, 2001:

$$\frac{\Delta f}{f} = -2.24 \times 10^{-13} \pm 0.02 \times 10^{-13}. \quad (93)$$

The quoted uncertainty is due to the combined uncertainties from the determination of the energy per unit mass before and after the orbit change. These uncertainties are due to neglecting tidal forces of the sun and moon, radiation pressure, and other non-gravitational forces.

Summary. We note that the values of semi-major axis reported by Epstein *et al.* [12] differ from the values obtained by averaging as outlined above, by 200–300 m. This difference arises because of the different methods of calculation. In the present calculation, an attempt was made to account for the effect of earth's quadrupole moment on the Keplerian orbit. It was not necessary to compute the orbit eccentricity. Agreement with measurement of the fractional frequency shift was only a few percent better than that obtained by differencing the maximum and minimum radii. This approximate treatment of the orbit makes no attempt to consider perturbations that are non-gravitational in nature, *e.g.*, solar radiation pressure. The work was an investigation of the approximate effect of earth's quadrupole moment on the GPS satellite orbits, for the purpose of (possibly) accurate calculations of the fractional frequency shifts that result from orbit changes.

As a general conclusion, the fractional frequency shift can be estimated to very good accuracy from the expression for the “factory frequency offset”.

$$\frac{\delta f}{f} = +\frac{3GM_E \delta a}{2c^2 a^2}. \quad (94)$$

10 Secondary Relativistic Effects

There are several additional significant relativistic effects that must be considered at the level of accuracy of a few cm (which corresponds to 100 picoseconds of delay). Many investigators are modelling systematic effects down to the millimeter level, so these effects, which currently are not sufficiently large to affect navigation, may have to be considered in the future.

Signal propagation delay. The Shapiro signal propagation delay may be easily derived in the standard way from the metric, Eq. (23), which incorporates the choice of coordinate time rate expressed by the presence of the term in Φ_0/c^2 . Setting $ds^2 = 0$ and solving for the increment of coordinate time along the path increment $d\sigma = \sqrt{dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2}$ gives

$$dt = \frac{1}{c} \left[1 - \frac{2V}{c^2} + \frac{\Phi_0}{c^2} \right] d\sigma. \quad (95)$$

The time delay is sufficiently small that quadrupole contributions to the potential (and to Φ_0) can be neglected. Integrating along the straight line path a distance l between the transmitter and receiver gives for the time delay

$$\Delta t_{\text{delay}} = \frac{\Phi_0}{c^2} \frac{l}{c} + \frac{2GM_E}{c^3} \ln \left[\frac{r_1 + r_2 + l}{r_1 + r_2 - l} \right], \quad (96)$$

where r_1 and r_2 are the distances of transmitter and receiver from earth's center. The second term is the usual expression for the Shapiro time delay. It is modified for GPS by a term of opposite sign (Φ_0 is negative), due to the choice of coordinate time rate, which tends to cancel the logarithm term. The net effect for a satellite to earth link is less than 2 cm and for most purposes can be neglected. One must keep in mind, however, that in the main term l/c , l is a coordinate distance and further small relativistic corrections are required to convert it to a proper distance.

Effect on geodetic distance. At the level of a few millimeters, spatial curvature effects should be considered. For example, using Eq. (23), the proper distance between a point at radius r_1 and another point at radius r_2 directly above the first is approximately

$$\int_{r_1}^{r_2} dr \left[1 + \frac{GM_E}{c^2 r} \right] = r_2 - r_1 + \frac{GM_E}{c^2} \ln \left(\frac{r_2}{r_1} \right). \quad (97)$$

The difference between proper distance and coordinate distance, and between the earth's surface and the radius of GPS satellites, is approximately $4.43 \ln(4.2) \text{ mm} \approx 6.3 \text{ mm}$. Effects of this order of magnitude would enter, for example, in the comparison of laser ranging to GPS satellites, with numerical calculations of satellite orbits based on relativistic equations of motion using coordinate times and coordinate distances.

Phase wrap-up. Transmitted signals from GPS satellites are right circularly polarized and thus have negative helicity. For a receiver at a fixed location, the electric field vector rotates counterclockwise, when observed facing into the arriving signal. Let the angular frequency of the signal be ω in an inertial frame, and suppose the receiver spins rapidly with angular frequency Ω which is parallel to the propagation direction of the signal. The antenna and signal electric field vector rotate in opposite directions and thus the received frequency will be $\omega + \Omega$. In GPS literature this is described in terms of an accumulation of phase called “phase wrap-up”. This effect has been known for a long time [17, 20, 21, 24], and has been experimentally measured with GPS receivers spinning at rotational rates as low as 8 cps. It is similar to an additional Doppler effect; it does not affect navigation if four signals are received simultaneously by the receiver as in Eqs. (1). This observed effect raises some interesting questions about transformations to rotating, spinning coordinate systems.

Effect of other solar system bodies. One set of effects that has been “rediscovered” many times are the redshifts due to other solar system bodies. The Principle of Equivalence implies that sufficiently near the earth, there can be no linear terms in the effective gravitational potential due to other solar system bodies, because the earth and its satellites are in free fall in the fields of all these other bodies. The net effect locally can only come from tidal potentials, the third terms in the Taylor expansions of such potentials about the origin of the local freely falling frame of reference. Such tidal potentials from the sun, at a distance r from earth, are of order $GM_{\odot}r^2/R^3$ where R is the earth-sun distance [8]. The gravitational frequency shift of GPS satellite clocks from such potentials is a few parts in 10^{16} and is currently neglected in the GPS.

11 Applications

The number of applications of GPS have been astonishing. It would take several paragraphs just to list them. Accurate positioning and timing, other than for military navigation, include synchronization of power line nodes for fault detection, communications, VLBI, navigation in deep space, tests of fundamental physics, measurements on pulsars, tests of gravity theories, vehicle tracking, search and rescue, surveying, mapping, and navigation of commercial aircraft, to name a few. These are too numerous to go into in much detail here, but some applications are worth mentioning. Civilian applications have overtaken military applications to the extent that SA was turned off in May of 2000.

The Nobel-prizewinning work of Joseph Taylor and his collaborators [16, 23] on the measurement of the rate of increase of the binary pulsar period depended on GPS receivers at the Arecibo observatory, for transferring UTC from the U.S. Naval Observatory and NIST to the local clock. Time standards around the world are compared using GPS in common-view; with this technique SA would cancel out, as well as do many sources of systematic errors such as ionospheric and tropospheric delays. Precise position information can assist in careful husbandry of natural resources, and animal and vehicle fleet tracking can result in improved efficiency. Precision agriculture makes use of GPS receivers in real-time application of pesticides or fertilizers, minimizing waste. Sunken vessels or underwater ruins with historically significant artifacts can be located using the GPS and archeologists can return again and again with precision to the same location. Monster ore trucks or earth-moving machines can be fitted with receivers and controlled remotely with minimal risk of collision or interference with other equipment. Disposable GPS receivers dropped through tropical storms transmit higher resolution measurements of temperature, humidity, pressure, and wind speed than can be obtained by any other method; these have led to improved understanding of how tropical storms intensify. Slight movements of bridges or buildings, in response to various loads, can be monitored in real time. Relative movements of remote parts of earth's crust can be accurately measured in a short time, contributing to better understanding of tectonic processes within the earth and, possibly, to future predictions of earthquakes. With the press of a button, a lost hiker can send a distress signal that includes the hikers' location.

These and many other creative applications of precise positioning and timing are leading to a rapid economic expansion of GPS products and services. Over 50 manufacturers produce more than 350 different GPS products for commercial, private, and military use. The number of receivers manufactured each year is in excess of two million, and different applications are continually being invented. Recent marketing studies predict that sales of GPS equipment and services will grow to over \$34 billion by 2006. Revenue for the European Galileo system is projected to be 10 billion Euros per year.

12 Conclusions

The GPS is a remarkable laboratory for applications of the concepts of special and general relativity. GPS is also valuable as an outstanding source of pedagogical examples. It is deserving of more scrutiny from relativity experts. It is particularly important to confirm that the basis for synchronization is on a firm conceptual foundation. A number of other agencies have expressed interest in establishing alternatives to the GPS, since this system is under the military control.

Plans are being made by several countries to put laser-cooled clock(s) having stabilities of $5 \times 10^{-14}/\sqrt{\tau}$ and accuracies of 1×10^{16} , on the International Space Station. This will open up additional possibilities for testing relativity as well as for making improvements in GPS and in other potential navigational satellite systems.

References

- [1] “IAU Resolutions adopted at the 24th General Assembly, Manchester, August 2000”, (2000), [Online HTML Page]: cited on 20 December 2002, http://syrtte.obspm.fr/IAU_resolutions/Resol-UA1.htm. 1
- [2] *NAVSTAR GPS Space Segment and Navigation User Interfaces*, (ARINC Research Corporation, Fountain Valley, CA, 1993), Report no. ICD-GPS-200C. 5, 9
- [3] *World Geodetic System 1984*, (National Imagery and Mapping Agency, 1994), Report no. 8350.2, Third Edition, Amendment 1. NIMA Stock No. DMATR83502WGS84, MSN 7643-01-402-0347. 2
- [4] Allan, D.W., Ashby, N., and Weiss, M., “Around-the-World Relativistic Sagnac Experiment”, *Science*, **228**, 69–70, (1985). 2
- [5] Alley, C., “Proper time experiments in gravitational fields with atomic clocks, aircraft, and laser light pulses”, in Meystre, P., and Scully, M. O., eds., *Quantum Optics, Experimental Gravity, and Measurement Theory*, Proc. NATO Advanced Study Institute on Quantum Optics and Experimental General Relativity, August 1981, Bad Windsheim, Germany, 363, (Plenum Press, New York, 1983). 5
- [6] Ashby, N., *An Earth-Based Coordinate Clock Network*, (U.S. Dept. of Commerce, U.S. Government Printing Office, Washington, D.C. 20402, 1975), Report no. 659 (NBS Technical Note). S.D. Catalog # C13:46:659. 2
- [7] Ashby, N., and Allan, D.W., “Practical Implications of Relativity for a Global Coordinate Time Scale”, *Radio Sci.*, **14**, 649–669, (1979). 4
- [8] Ashby, N., and Bertotti, B., “Relativistic Effects in Local Inertial Frames”, *Phys. Rev. D*, **34**, 2246–2258, (1986). 10
- [9] Ashby, N., and Spilker Jr., J.J., “Introduction to relativistic effects on the Global Positioning System”, in Parkinson, B.W., and Spilker Jr., J.J., eds., *Global Positioning System: Theory and Applications*, volume 1, chapter 18, 623–697, (American Institute of Aeronautics and Astronautics, Inc., Washington, D.C., 1996). 1, 9
- [10] Ashby, N., and Weiss, M., *Global Positioning Receivers and Relativity*, (U.S. Government Printing Office, Washington, D.C. 20402, 1999), Report no. 1385 (NIST Technical Note). 2
- [11] Buisson, J.A., Easton, R.L., and McCaskill, T.B., “Initial Results of the NAVSTAR GPS NTS-2 Satellite”, in Rueger, L. et al., ed., *Proceedings of the 9th Annual Precise Time and Time Interval Applications and Planning Meeting*, 177–200, (Technical Information and Administrative Support Division, Goddard Space Flight Center, Greenbelt, MD, 1977). 5

- [12] Epstein, M., Stoll, E., and Fine, J., “Study of SVN43 clock from 06/18/00 to 12/30/00”, in Breakiron, L., ed., *Proceedings of the 33rd Annual Precise Time and Time Interval Systems and Applications Meeting, Long Beach, Nov. 2001*, (U.S. Naval Observatory, Washington, D.C., 2001). 9, 9, 9, 9, 9
- [13] Fitzpatrick, P., *The Principles of Celestial Mechanics*, (Academic Press, New York, 1970). 5, 9, 9, 9, 9
- [14] Fliegel, H. F., and DiEsposti, R. S., *GPS and Relativity: an Engineering Overview*, (The Aerospace Corporation, 1996), Report no. ATR-97(3389)-1 (Aerospace Corporation Report). 6
- [15] Garland, G. D., *The Earth’s Shape and Gravity*, (Pergamon Press, New York, 1965). 1
- [16] Hulse, R.A., and Taylor, J.H., “Discovery of a pulsar in a binary system”, *Astrophys. J.*, **195**, L51–53, (1975). 11
- [17] Kraus, J. D., *Antennas*, (McGraw-Hill, 1988), 2nd edition. Reprinted by Cygnus-Quasar Books, Powell, Ohio. 10
- [18] Lambeck, K., “Geophysical Geodesy: The Slow Deformations of the Earth”, 13–18, Oxford Science Publications, (Oxford University Press, Oxford, 1988). 1
- [19] Malys, S., and Slater, J., “Maintenance and Enhancement of the World Geodetic System 1984”, in *Proceedings of the 7th International Technical Meeting of The Satellite Division of The Institute of Navigation, Sept. 20–23, Salt Lake City, UT (ION GPS-94)*, 17–24. The Institute of Navigation, (1994). 2
- [20] Mashhoon, B., “Nonlocal Theory of Accelerated Observers”, *Phys. Rev. A*, **47**, 4498–4501, (1993). 10
- [21] Mashhoon, B., “On the Coupling of Intrinsic Spin with the Rotation of the Earth”, *Phys. Lett. A*, **198**, 9–13, (1995). 10
- [22] Parkinson, B.W., and Spilker Jr., J.J., “Overview of GPS Operation and Design”, in Parkinson, B.W., and Spilker Jr., J.J., eds., *Global Positioning System: Theory and Applications*, volume 1, chapter 2, 33, (American Institute of Aeronautics and Astronautics, Inc., Washington, D.C., 1996). 6
- [23] Taylor, J.H., “Binary Pulsars and Relativistic Gravity”, *Rev. Mod. Phys.*, **66**, 711–719, (1994). 11
- [24] Tetewsky, A.K., and Mullen, F.E., “Effects of Platform Rotation on GPS with Implications for GPS Simulators”, in *Proceedings of the 9th International Technical Meeting of the Satellite Division of the Institute of Navigation, Sept. 17–20, 1996, Kansas City, MO (ION GPS-96)*, 1917–1925. The Institute of Navigation, (1996). 10