24. Relativity in GNSS

Neil Ashby

Global navigation satellite systems (GNSS) use accurate, stable atomic clocks in satellites and on the ground to provide world-wide position, velocity, and time to millions of users. Orbiting clocks have gravitational and motional frequency shifts that are so large that, without carefully accounting for numerous relativistic effects, the systems would not work. The basis for navigation using GNSS, founded on special and general relativity, includes relativistic principles, concepts and effects such as the constancy of the speed of light, relativity of synchronization, coordinate time, proper time, time dilation, the Sagnac effect, the weak equivalence principle, and gravitational frequency shifts. Additional small relativistic effects such as the coordinate slowing of light speed and the effects of tidal potentials from the moon and the sun may need to be accounted for in the future. Examples of new navigation systems that are being developed and deployed are the European GALILEO system and the Chinese BEIDOU system; these will greatly widen the impact of GNSS. This chapter discusses applications of relativistic concepts in GNSS.

Since the first deployment of high-performance atomic clocks in satellites in 1977, position, navigation, and timing have been revolutionized world-wide. The United States' global positioning system (GPS), the Russian global navigation satellite system (GLONASS – globalnaya navigatsionnaya sputnikovaya sistema), the European GALILEO system, and China's BEIDOU system will soon provide 100 or more satellites with synchronized clocks in precisely determined orbits. Each system consists of approximately 30 satellites, capable of transmitting messages that enable a receiver to accurately compute its position, velocity, and time anywhere near earth's surface. There are also numerous augmentation systems designed to provide improved reliability and accuracy. Examples are the US's WAAS (wide area augmentation system), which uses geosynchronous satellites to broadcast GPS-like signals over the continental United States, and Japan's QZSS system that uses satellites in highly eccentric orbits, enabling them to spend considerable time directly over an area of particular interest. These systems together are generally referred to as GNSS.

A vast infrastructure supports these systems: world-wide networks of receivers and organizations to monitor and estimate the satellite orbits and clocks; ensembles of high-performance clocks on the ground to provide time references; industries to design, manufacture, and launch the satellites; and hundreds of millions of users with receivers of varying degrees of complexity and expense. The GPS infrastructure has been adequately described elsewhere [24.1].

The remarkable positioning precision achieved by GNSS is due to careful accounting for a number of systematic effects that would otherwise greatly degrade the results and eventually render the system useless. Among these effects are signal delays due to water vapor in the troposphere, free electrons in the ionosphere, and reflections of signals from surfaces near the receiver antenna. Unless relativistic concepts and effects on clocks and radio signals in the GNSS are taken into account, the systems will not work. This article discusses the fundamental principles of special and general relativity that provide the basis for positioning in the GNSS. The principle of equivalence is discussed in Sect. 24.2, where it is shown that to a first approximation, gravitational potentials due to the sun and the moon can be neglected in the GNSS. Relative motions of clocks and the rotation of the earth leads to the discussion of coordinate time and the Sagnac effect in Sect. 24.3. In Sect. 24.4 we discuss international atomic time (TAI) and universal coordinated time (UTC). Sections Sect. 24.5 through Sect. 24.9 discuss relativistic effects on ground-based clocks and orbiting clocks and how such effects are accounted for. Additional effects that are currently neglected are described in Sect. 24.10.

24.1 The Principle of Equivalence

The *weak* equivalence principle is based on the observed universality of free fall, namely that all objects fall with equal accelerations in a given gravitational field, independent of their internal structure, mass, or composition. Thus in a freely falling laboratory of sufficiently small extent, no experiment performed locally – entirely within the laboratory – can tell that the laboratory is in free fall. Although this has been tested only to a certain, very high level of precision [24.2], it means that even if there is no gravitational field due to nearby masses, then in a uniformly accelerating laboratory an induced gravitational field will appear that can in no way, by local measurements only, be distinguished from a real gravitational field.

Some have been tempted to think that clocks in satellites, which are momentarily on the side of the earth nearest the sun, are affected more by the sun than satellites on the side of earth away from the sun; one implication that has been put forward numerous times is that clocks in satellites nearer the sun suffer a greater shift in frequency toward the red than do clocks on the opposite side of the earth. By the principle of equivalence, however, this picture is erroneous.

The earth and its satellites are in free fall about the sun, moon, and other solar system bodies. Locally, the gravitational field due to external bodies causes acceleration, which in turn induces an equal but opposite fictitious gravitational field; these can be superimposed and they cancel to high precision near earth's center of mass. Let the total gravitational potential in the neighborhood of the earth be denoted by $\Phi(\boldsymbol{r})$; it will be the sum of earth's potential, $V(\boldsymbol{r})$, plus the potential due to external sources, $\phi_{\text{ext}}(\boldsymbol{r})$

$$\Phi(\boldsymbol{r}) = V(\boldsymbol{r}) + \phi_{\text{ext}}(\boldsymbol{r}) , \tag{24.1}$$

where $\boldsymbol{r} = \{x^1, x^2, x^3\}$ is a vector from the center of mass of the earth to the point of observation. We take the origin of spatial coordinates to be earth's center of mass. The distance $r = |\boldsymbol{r}|$ is small compared to the distance to any external source, so we may imagine a series expansion of the external potential about earth's center of mass

$$\begin{aligned}\Phi(\boldsymbol{r}) &= \phi(\boldsymbol{r}) + \phi_{\text{ext}}(0) \\ &\quad + \sum_{i=1}^{3} x^i \frac{\partial \phi_{\text{ext}}}{\partial x^i}\bigg|_0 + \frac{1}{2}\sum_{i,j=1}^{3} x^i x^j \frac{\partial^2 \phi_{\text{ext}}}{\partial x^i \partial x^j}\bigg|_0 \\ &\quad + \ldots .\end{aligned} \tag{24.2}$$

The term $\phi_{\text{ext}}(0)$ represents a constant potential everywhere near the earth and affects all physical objects in the same way. It cannot be detected and thus can be ignored. The linear terms on the second line of (24.2) represent the strength of the gravitational field due to external sources and are canceled by the induced gravitational field due to the acceleration. This is not easy to prove from first principles but proofs can be found in the literature [24.3–5]. Evidence for this result is that the linear term would exert a huge effect on the oceans, whereas it is only the last term in (24.2) that gives rise to the ocean tides. For most purposes in the GNSS tidal effects on clocks are small and can at first be neglected. The tidal effects will be discussed further in Sect. 24.9. We conclude that for GNSS, to a high degree of approximation the only gravitational potential of significance is that of the earth itself. Although the earth and its satellites fall freely in the gravitational fields of external sources, one can introduce coordinate axes with origin at earth's center of mass and axes pointing toward distant references in the cosmos; this defines a reference system which is locally very nearly inertial. In such a system clocks can be synchronized using constancy of the speed of light.

24.2 Navigation Principles in the GNSS

The principles of position determination and time transfer in the GNSS can be very simply stated. Let there be four synchronized atomic clocks which transmit sharply defined pulses from the positions $\boldsymbol{r}_j$ at times t_j, with $j = 1, 2, 3, 4$ an index labeling the different transmission events.

Then from the principle of the constancy of the speed of light

$$c^2(t-t_j)^2 = |\boldsymbol{r}-\boldsymbol{r}_j|^2\,, \quad j=1,2,3,4\,, \tag{24.3}$$

where the defined value of c is exactly 299 792 458 m/s. These four equations can be solved for the unknown space-time coordinates of the reception event, $\{\boldsymbol{r}, t\}$. Hence the principle of the constancy of c finds application as the fundamental concept on which navigation and timing in the GNSS is based. Obviously, it is necessary to specify carefully the reference frame in which the transmitter clocks are synchronized, so that (24.3) is valid.

Equation (24.3) is nonlinear. Typically solutions are obtained by linearizing, solving approximately, and then iterating until a solution converges. For example, if one guesses that the solution is $\boldsymbol{r} = \boldsymbol{r}_0 + \delta(\boldsymbol{r}), ct = ct_0 + \delta(ct)$, where the corrections $\delta(\boldsymbol{r})$ and $\delta(ct)$ are small, then linearizing the navigation equations gives

$$\boldsymbol{N}_j \cdot \delta(\boldsymbol{r}) - \delta(ct) = c(t_0 - t_j) - |\boldsymbol{r}_0 - \boldsymbol{r}_j|\,, \tag{24.4}$$

where $\boldsymbol{N}_j$ is a unit vector from the j-th satellite to the assumed receiver position. Four such equations can be written in matrix form and the matrix equation can be solved for the corrections; iteration of the calculation usually converges very rapidly because the distances between receiver and satellites are large compared to the distance from earth's center to the receiver.

Equation (24.4) also allows one to estimate position uncertainties arising from uncertainties in determining the propagation time intervals or from poor satellite geometry. For example, suppose a receiver is at the geometric center of a tetrahedral satellite configuration and that timing errors from the satellites are uncorrelated and are each 10 ns (1 ns $= 10^{-9}$ s); 10 ns corresponds to a position error of 3 m in each direction resulting in an estimated position which is within a sphere of radius 4.7 m. In real navigation situations such ideal tetrahedral symmetry cannot be achieved since the earth's presence forces the received signals to come from somewhat less than 2π steradians of the sky above. The position error then crucially depends on the independence of the vectors $\boldsymbol{N}_j$; if these vectors should all lie close to some plane then the position uncertainty can be many times larger. Thus, the navigation equations play an important role in design of the satellite configuration so that such errors are minimized.

Signals transmitted to users from the satellites are right circularly polarized. Usually information is transmitted by encoding the high frequency carriers with phase reversals. The timing signals in question can then 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; these are relativistically invariant events and, therefore, provide relatively moving observers with sequences of events that they can agree on in principle.

24.3 Rotation and the Sagnac Effect

Almost all users of GNSS 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 in the GPS 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, $\omega_{\mathrm{E}} = 7.292115 \times 10^{-5}\,\mathrm{rad\,s^{-1}}$. This reference frame is designated by the symbol WGS-84; the station coordinates used to define this system have been updated several times since 1984 [24.6–8]. The latest realization is termed WGS-84(G1150) and is generally assumed to be identical to the International Terrestrial Reference Frame ITRF00 [24.8]. The differences among these frames are only a few centimeters. Other GNSS systems use their own earth-fixed reference systems. The Galileo terrestrial reference frame (GTRF) is an independent realization of the International Terrestrial Reference System (ITRS) established by the Central Bureau of the International Earth Rotation Service (IERS). For discussions of relativity, the particular choice of ECEF frame is immaterial. Also, the fact that the earth truly rotates about a slightly different axis with a variable rotation rate has little consequence for relativity and will not be discussed here. We shall simply regard the ECEF frame of the appropriate GNSS system as closely related to, or determined by, the ITRF established by the International Bureau of Weights and Measures (BIPM).

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 (24.3), the receiver must generally perform a different rotation for each measurement made, into some common inertial frame, so that (24.3) 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 [24.9] discusses these issues in considerable detail.

Although the ECEF frame is of primary interest for navigation, it is simpler to describe many physical processes (such as electromagnetic wave propagation) in an inertial reference frame. Certainly, inertial reference frames are needed to express (24.3), whereas it would lead to serious error to assert (24.3) 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.

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. Ignoring gravitational potentials for the moment, the metric in an inertial frame in cylindrical coordinates is

$$-\mathrm{d}s^2 = -(c\,\mathrm{d}t)^2 + \mathrm{d}r^2 + r^2\,\mathrm{d}\phi^2 + \mathrm{d}z^2\,, \qquad (24.5)$$

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

$$\begin{aligned} t &= t'\,, & r &= r'\,,\\ \phi &= \phi' + \omega_{\mathrm{E}} t'\,, & z &= z'\,. \end{aligned} \qquad (24.6)$$

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

$$\begin{aligned} -\mathrm{d}s^2 = &-\left(1 - \frac{\omega_{\mathrm{E}}^2 r'^2}{c^2}\right)(c\,\mathrm{d}t')^2 \\ &+ 2\omega_{\mathrm{E}} r'^2\,\mathrm{d}\phi'\,\mathrm{d}t' + (\mathrm{d}\sigma')^2\,, \end{aligned} \qquad (24.7)$$

where the abbreviated expression $(\mathrm{d}\sigma')^2 = (\mathrm{d}r')^2 + (r'\,\mathrm{d}\phi')^2 + (\mathrm{d}z')^2$ for the square of the coordinate distance has been used.

The time transformation $t = t'$ in (24.6) is deceivingly 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 GNSS.

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 $\mathrm{d}s^2 = 0$ in (24.7). Also, it is sufficient for this discussion to keep only terms of first order in the small parameter $\omega_{\mathrm{E}} r'/c$. Then

$$(c\,\mathrm{d}t')^2 - \frac{2\omega_{\mathrm{E}} r'^2\,\mathrm{d}\phi'(c\,\mathrm{d}t')}{c} - (\mathrm{d}\sigma')^2 = 0\,, \qquad (24.8)$$

and solving for $(c\,\mathrm{d}t')$,

$$c\,\mathrm{d}t' = \mathrm{d}\sigma' + \frac{\omega_{\mathrm{E}} r'^2\,\mathrm{d}\phi'}{c}\,. \qquad (24.9)$$

The quantity $r'^2\,\mathrm{d}\phi'/2$ is just the infinitesimal area $\mathrm{d}A'_z$ 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

$$\int_{\text{path}} \mathrm{d}t' = \int_{\text{path}} \frac{\mathrm{d}\sigma'}{c} + \frac{2\omega_{\mathrm{E}}}{c^2} \int_{\text{path}} \mathrm{d}A'_z \quad [\text{light}] \,. \tag{24.10}$$

Observers fixed on the earth, who were unaware of earth rotation, would use just $\int \mathrm{d}\sigma'/c$ 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 which follows earth's equator eastward around the globe. For earth, $2\omega_{\mathrm{E}}/c^2 = 1.6227 \times 10^{-21}$ s/m^2 and the equatorial radius is $a_1 = 6\,378\,137$ m, so the area is $\pi a_1^2 = 1.27802 \times 10^{14}$ m^2 . Thus the last term in (24.10) is

$$\frac{2\omega_{\mathrm{E}}}{c^2} \int_{\text{path}} \mathrm{d}A'_z = 207.4\,\text{ns} \,. \tag{24.11}$$

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. 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 $\int \mathrm{d}\sigma'/c$ and thus leads to a significant error.

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 (24.7), the proper time increment $\mathrm{d}\tau$ on the moving clock is given by

$$\begin{aligned}(\mathrm{d}\tau)^2 &= \left(\frac{\mathrm{d}s}{c}\right)^2 \\ &= \mathrm{d}t'^2 \left[1 - \left(\frac{\omega_{\mathrm{E}} r'}{c}\right)^2 - \frac{2\omega_{\mathrm{E}} r'^2 \,\mathrm{d}\phi'}{c^2\,\mathrm{d}t'} - \left(\frac{\mathrm{d}\sigma'}{c\,\mathrm{d}t'}\right)^2\right].\end{aligned} \tag{24.12}$$

For a slowly moving clock $(\mathrm{d}\sigma'/c\,\mathrm{d}t')^2 << 1$, so the last term in brackets in (24.12) can be neglected. Also, keeping only first-order terms in the small quantity $\omega_{\mathrm{E}} r'/c$

$$\mathrm{d}\tau = \mathrm{d}t' - \frac{\omega_{\mathrm{E}} r'^2 \,\mathrm{d}\phi'}{c^2} \,, \tag{24.13}$$

which leads to

$$\int_{\text{path}} \mathrm{d}t' = \int_{\text{path}} \mathrm{d}\tau + \frac{2\omega_e}{c^2} \int_{\text{path}} \mathrm{d}A'_z \quad [\text{portable clock}] \,. \tag{24.14}$$

This should be compared with (24.10). 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.

Equations (24.10) and (24.14) 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 $+2\omega_{\mathrm{E}} \int_{\text{path}} \mathrm{d}A'_z/c^2$ 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 comoving 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 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 GNSS 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 [24.10].

Satellite clocks 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.

24.4 Coordinate Time and TAI

For GNSS 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, we next 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, nonrotating, 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

$$-\mathrm{d}s^2 = -\left(1+\frac{2V}{c^2}\right)(c\,\mathrm{d}t)^2 + \left(1-\frac{2V}{c^2}\right) \times (\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2)\,, \tag{24.15}$$

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

$$V = -\frac{GM_\mathrm{E}}{r}\left[1 - J_2\left(\frac{a_1}{r}\right)^2 P_2(\cos\theta)\right]. \tag{24.16}$$

In (24.16), $GM_\mathrm{E} = 3.986004418\times10^{14}\,\mathrm{m^3\,s^{-2}}$ is the product of earth's mass times the Newtonian gravitational constant, $J_2 = 1.0826300\times10^{-3}$ is earth's quadrupole moment coefficient, and $a_1 = 6.3781370\times10^6$ is earth's equatorial radius. (WGS-84(G1150) values of these constants are used in this article [24.8].) 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 (24.15), it is an adequate approximation to retain only terms of first order in the small quantity V/c^2. Higher multipole moment contributions to (24.16) have very small effect on relativity in GNSS.

One additional expression for the invariant interval is needed, the transformation of (24.16) to a rotating, ECEF coordinate system by means of transformations equivalent to (24.6). The transformations for spherical polar coordinates are

$$t = t'\,,\quad r = r'\,,\quad \theta = \theta'\,,\quad \phi = \phi' + \omega_\mathrm{E} t'\,. \tag{24.17}$$

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

$$-\mathrm{d}s^2 = -\left[1+\frac{2V}{c^2}-\left(\frac{\omega_\mathrm{E} r'\sin\theta'}{c}\right)^2\right](c\,\mathrm{d}t')^2 + 2\omega_\mathrm{E} r'^2\sin^2\theta'\,\mathrm{d}\phi'\,\mathrm{d}t' + \left(1-\frac{2V}{c^2}\right) \times (\mathrm{d}r'^2 + r'^2\,\mathrm{d}\theta'^2 + r'^2\sin^2\theta'\,\mathrm{d}\phi'^2)\,. \tag{24.18}$$

To the order of the calculation, this result is a simple superposition of the metric, (24.15), with the corrections due to rotation expressed in (24.17). The metric tensor coefficient g'_{00} in the rotating frame is

$$g'_{00} = -\left[1+\frac{2V}{c^2}-\left(\frac{\omega_e r'\sin\theta'}{c}\right)^2\right] \equiv -\left(1+\frac{2\Phi}{c^2}\right), \tag{24.19}$$

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.

24.4.1 The Earth's Geoid

In (24.16) and (24.17), the rate of coordinate time is determined by atomic clocks at rest at infinity. The rate of coordinate time used in GNSS, however, is closely related to international atomic time (TAI), which is a time scale computed by the (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 a time scale that 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 which contribute to TAI and UTC generally maintain their own time scales. For example, the time scale of the US 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 of UTC(USNO). Usually this steering is so successful that the difference between GPS time and UTC(USNO) is of order 10 ns. Receiver 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 GNSS are not at infinity, We need to 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, (24.18) in the ECEF is relevant. For a clock at rest on earth, (24.18) reduces to

$$-\mathrm{d}s^2 = -\left(1 + \frac{2V}{c^2} - \frac{\omega_e^2 r'^2 \sin^2\theta'}{c^2}\right)(c\,\mathrm{d}t')^2 \,. \tag{24.20}$$

with the potential V given by (24.16).

This equation determines the radius r' of the effective equipotential geoid surface as a function of polar angle θ'. The numerical value of Φ_0 at the geoid can be determined at the equator where $\theta' = \pi/2$ and $r' = a_1$. This gives

$$\begin{aligned}\frac{\Phi_0}{c^2} &= -\frac{GM_\mathrm{E}}{a_1c^2} - \frac{GM_\mathrm{E}J_2}{2a_1c^2} - \frac{\omega_\mathrm{E}^2a_1^2}{2c^2}\\ &= -6.95348\times10^{-10}\\ &\quad -3.764\times10^{-13} - 1.203\times10^{-12}\\ &= -6.96927\times10^{-10}\,.\end{aligned} \tag{24.21}$$

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 [24.11] a *terrestrial time* scale (TT) has been defined by defining the value $\Phi_0/c^2 = 6.969290134\times10^{-10}$. Equation (24.21) agrees with this definition to within the accuracy needed for the GNSS.

From (24.18), for clocks on the geoid,

$$\mathrm{d}\tau = \frac{\mathrm{d}s}{c} = \mathrm{d}t'\left(1 + \frac{\Phi_0}{c^2}\right)\,. \tag{24.22}$$

Clocks at rest on the rotating geoid run slow compared to clocks at rest at infinity by about seven parts in 10^{10}. 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 (24.19) for r' in terms of θ'. The first few terms in a power series in the variable $x' = \sin\theta'$ can be expressed as

$$\begin{aligned}r' &= 6\,356\,742.025 + 21\,353.642x'^2 + 39.832x'^4\\ &\quad + 0.798x'^6 + 0.003x'^8\mathrm{m}\,.\end{aligned} \tag{24.23}$$

This treatment of the gravitational field of the oblate earth is limited by the simple model of the gravitational field. Actually (24.23) estimates 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 [24.12–14]. 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 (24.23) by as much as 100 m, the effects of such terms on timing in GNSS are small. Incorporating up to 20 higher zonal harmonics in a calculation Φ_0 affects the value 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 (24.22), Φ_0 is a constant. On the left-hand side of (24.22), $\mathrm{d}\tau$ is the increment of proper time elapsed on a standard clock at rest, in terms of the elapsed coordinate time $\mathrm{d}t$. 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 here we 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 goes 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. Equation (24.15) defines the rate of coordinate time in terms of the rate of standard clocks at rest at infinity. What is needed instead is to define the rate of coordinate time by standard clocks at rest on earth's geoid. Therefore, we define a new coordinate time t'' by means of a constant rate change

$$t'' = (1 + \Phi_0/c^2)t' = (1 + \Phi_0/c^2)t\,. \tag{24.24}$$

The correction is about seven parts in 10^{10} (see (24.21)).

When this time scale change is made, the metric of (24.18) in the earth-fixed rotating frame becomes

$$\begin{aligned}-\mathrm{d}s^2 = &-\left(1 + \frac{2(\Phi - \Phi_0)}{c^2}\right)(c\,\mathrm{d}t'')^2 \\ &+ 2\omega_\mathrm{E} r'^2 \sin^2\theta'\,\mathrm{d}\phi'\,\mathrm{d}t'' \\ &+ \left(1 - \frac{2V}{c^2}\right) \\ &\times (\mathrm{d}r'^2 + r'^2\,\mathrm{d}\theta'^2 + r'^2\sin^2\theta'\,\mathrm{d}\phi'^2)\,.\end{aligned} \tag{24.25}$$

where only terms of order c^{-2} have been retained. Whether $\mathrm{d}t'$ or $\mathrm{d}t''$ is used in the Sagnac cross term makes no difference since the Sagnac term is very small anyway. The same time scale change in the nonrotating ECI metric, (24.15), gives

$$\begin{aligned}-\mathrm{d}s^2 = &-\left(1 + \frac{2(V - \Phi_0)}{c^2}\right)(c\,\mathrm{d}t'')^2 \\ &+ \left(1 - \frac{2V}{c^2}\right) \\ &\times (\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2)\,.\end{aligned} \tag{24.26}$$

Equations (24.25) and (24.26) 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.

24.5 The Realization of Coordinate Time

We are now able to address the real problem of clock synchronization within GNSS. In the remainder of this paper we drop the primes on t'' and just use the symbol t, with the understanding that unit of this time is referenced to one of the realizations of UTC on the rotating geoid, but with synchronization established in an underlying, locally inertial, reference frame. The metric (24.26) will henceforth be written as

$$\begin{aligned}-\mathrm{d}s^2 = &-\left(1 + \frac{2(V - \Phi_0)}{c^2}\right)(c\,\mathrm{d}t)^2 \\ &+ \left(1 - \frac{2V}{c^2}\right) \\ &\times (\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2)\,.\end{aligned} \tag{24.27}$$

The difference $(V - \Phi_0)$ that appears in the first term of (24.27) arises because in the underlying earth-centered, locally inertial (ECI) coordinate system in which the equation is expressed, the unit of time is determined by moving clocks in a spatially dependent gravitational field.

Obviously (24.27) 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 GNSS clocks cannot simply be 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 (24.27) is valid in a coordinate patch large enough to cover the earth and the GNSS satellite constellations. Equation (24.27) 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 (24.27) represents a fairly simple and reasonable choice of gauge.) It is natural, therefore, to propose that the coordinate time variable t of (24.27) and (24.25) 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 (24.26) for $\mathrm{d}t$ as follows. First factor out $(c\,\mathrm{d}t)^2$ from all terms on the right-hand side

$$\begin{aligned}-\,\mathrm{d}s^2 &= -\Bigg[1+\frac{2(V-\Phi_0)}{c^2} \\ &\quad -\left(1-\frac{2V}{c^2}\right)\frac{\mathrm{d}r^2+r^2\,\mathrm{d}\theta^2+r^2\sin^2\theta\,\mathrm{d}\phi^2)}{(c\,\mathrm{d}t)^2}\Bigg] \\ &\quad \times (c\,\mathrm{d}t)^2\,. \end{aligned} \tag{24.28}$$

Simplify by writing the velocity in the ECI coordinate system as

$$v^2=\frac{\mathrm{d}r^2+r^2\,\mathrm{d}\theta^2+r^2\sin^2\theta\,\mathrm{d}\phi^2}{\mathrm{d}t^2}\,. \tag{24.29}$$

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

$$\mathrm{d}\tau=\frac{\mathrm{d}s}{c}=\left[1+\frac{(V-\Phi_0)}{c^2}-\frac{v^2}{2c^2}\right]\mathrm{d}t\,. \tag{24.30}$$

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

$$\int_{\text{path}}\mathrm{d}t=\int_{\text{path}}\mathrm{d}\tau\left[1-\frac{(V-\Phi_0)}{c^2}+\frac{v^2}{2c^2}\right]. \tag{24.31}$$

The relativistic effect on the clock, given in (24.30), is thus corrected by (24.31).

Suppose for a moment there were no gravitational fields. Then one could picture an underlying nonrotating 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 nonrotating 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 (24.31). 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 which would be measured by standard clocks at rest in the local inertial frame [24.15].

When the gravitational field due to the earth is considered, the picture is only a little more complicated. There still exists a coordinate time which can be found by computing a correction for gravitational redshift, given by the first correction term in (24.31).

24.6 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. The term Φ_0 in (24.30) includes the scale correction needed in order to use clocks at rest on the earth's surface as references. Earth's quadrupole contributes to Φ_0 in the term $-GM_\mathrm{E}J_2/2a_1$ in (24.21); there it contributes a fractional rate correction of -3.76×10^{-13}. This effect must be accounted for in GNSS. 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 was neglected. This will be discussed in a later section, for the present we only note that earth's quadrupole potential effect on orbiting GNSS clocks is only about one part in 10^{14}.

24.6.1 Satellite Orbits

Let us assume that the satellites move along Keplerian orbits. This is a good approximation for GNSS satellites, but poor if the satellites are at low altitude. This assumption yields relations with which to simplify (24.31). Since the quadrupole (and higher multipole) parts of the earth's potential are neglected, in (24.31) the potential is $V = -GM_{\mathrm{E}}/r$. 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: [24.16] the distance r from the center of the earth to the satellite in ECI coordinates is

$$r = a(1-e^2)/(1+e\cos f)\,. \tag{24.32}$$

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

$$\cos f = \frac{\cos E - e}{1 - e\cos E}\,, \tag{24.33}$$

$$\sin f = \sqrt{1-e^2}\,\frac{\sin E}{1 - e\cos E}\,. \tag{24.34}$$

Then another way to write the radial distance r is

$$r = a(1 - e\cos E)\,. \tag{24.35}$$

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

$$E - e\sin E = \sqrt{\frac{GM_{\mathrm{E}}}{a^3}}\,(t - t_{\mathrm{p}})\,, \tag{24.36}$$

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

$$\frac{1}{2}v^2 - \frac{GM_{\mathrm{E}}}{r} = -\frac{GM_{\mathrm{E}}}{2a}\,. \tag{24.37}$$

Inserting (24.37) for v^2 into (24.31) results in the following expression for the elapsed coordinate time on the satellite clock

$$\Delta t = \int_{\text{path}} \mathrm{d}\tau \times \left[1 + \frac{3GM_{\mathrm{E}}}{2ac^2} + \frac{\Phi_0}{c^2} - \frac{2GM_{\mathrm{E}}}{c^2}\left(\frac{1}{a} - \frac{1}{r}\right)\right]. \tag{24.38}$$

The first two constant rate correction terms in (24.38) for GPS have the values

$$\begin{aligned}\frac{3GM_{\mathrm{E}}}{2ac^2} + \frac{\Phi_0}{c^2} &= +2.5046\times 10^{-10} \\ &\quad -6.9693\times 10^{-10} \\ &= -4.4647\times 10^{-10}\,.\end{aligned} \tag{24.39}$$

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

$$\begin{aligned}&[1 - 4.4647\times 10^{-10}] \times 10.23\,\mathrm{MHz} \\ &= 10.22999999543\,\mathrm{MHz}\,.\end{aligned} \tag{24.40}$$

This adjustment is accomplished on the ground before the clock is placed in orbit. Five sources of relativistic effects contribute to this frequency offset. This effect is formally incorporated into the GPS specifications [24.17] and into GLONASS [24.18] but is not mentioned in the formal GALILEO signal-in-space specifications [24.19].

For GNSS systems other than GPS, typically some choice is made concerning the nominal period required for the satellite's ground track to repeat. For GLONASS, the satellite periods are 16/17 of the GPS satellite periods, while for GALILEO, the ground track repeats after 17 orbits, which takes 10 days. For BEIDOU it appears that the satellites in medium earth orbit (MEO) will have repeating ground tracks after 13 orbits in 10 days. Table 24.1 gives the nominal semimajor axes and the fractional frequency offsets for several of the systems.

The purpose of this frequency offset is to make corrections applied by the receiver smaller, so the job of the receiver is easier. Typically navigation messages from the satellites contain three coefficients that enable

Table 24.1 Nominal values of SV clock frequency offsets

GNSS system	a (km)	$10^{10} \times \Delta f/f$
GPS	26 562.76	−4.46473
GLONASS	25 509.64	−4.36144
GALILEO	29 601.31	−4.72191
BEIDOU (MEO)	27 910.20	−4.58538
Geosynchronous	42 164.17	−5.39151

the receiver to make corrections for satellite clock errors. These coefficients are denoted by a_0, a_1, and a_2; a_0 is a time or synchronization error correction, a_1 is a frequency correction, and a_2 is a frequency drift correction. The coefficient a_2 is seldom used. Although it is quite possible to implement a system in which this *factory frequency offset* is not applied before launch, the transmitted navigation messages would have to transmit a much larger a_1 coefficient, in which the first few bits are always the same. This would be wasteful of resources and would limit the number of bits available for real variations in the actual frequency offsets.

Figure 24.1 shows a histogram of 271 values of the a_1 coefficient transmitted by the GLONASS satellites, sampled from the GLONASS broadcast ephemeris at the beginning of each year for the last 7 years. The average of this sample is very nearly zero, with an RMS variation of about 1.6×10^{-12}. In an ideal world this number would be zero. Thus for GLONASS the frequency offsets achieved are within about 4% of the desired value.

Small frequency shifts can arise from clock drift, launch vibrations, environmental changes, and other unavoidable effects such as the inability to launch the satellite into an orbit with precisely the desired semi-major axis. Because of such effects, it is difficult to use GNSS clocks to measure relativistic frequency shifts.

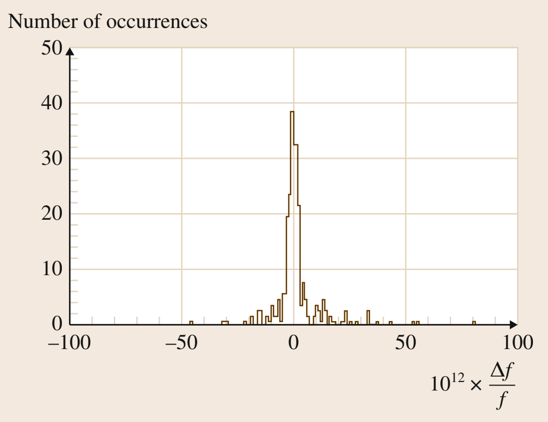

Fig. 24.1 Histogram of transmitted fractional frequency shift corrections for GLONASS. The horizontal axis is in units of 10^{-12}

24.6.2 The Eccentricity Correction

The last term in (24.38) may be integrated exactly by using the following expression for the rate of change of eccentric anomaly with time, which follows by differentiating (24.36)

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\sqrt{GM_\mathrm{E}/a^3}}{1 - e\cos E}\,. \tag{24.41}$$

A relativistic correction is being computed, so $\mathrm{d}s/c \simeq \mathrm{d}t$ and

$$\begin{aligned} &\int \left[\frac{2GM_\mathrm{E}}{c^2}\left(\frac{1}{r} - \frac{1}{a}\right)\right]\frac{\mathrm{d}s}{c} \\ &\simeq \frac{2GM_\mathrm{E}}{c^2}\int (1/r - 1/a)\,\mathrm{d}t \\ &= \frac{2GM_\mathrm{E}}{ac^2}\int \mathrm{d}t\left(\frac{e\cos E}{1 - e\cos E}\right) \\ &= \frac{2\sqrt{GM_\mathrm{E}a}}{c^2}e\,(\sin E - \sin E_0) \\ &= +\frac{2\sqrt{GM_\mathrm{E}a}}{c^2}e\sin E + \text{constant}\,. \end{aligned} \tag{24.42}$$

The constant of integration in (24.42) can be dropped since this term is lumped with other clock offset effects in the process of estimating the clock correction. The net correction for clock offset due to relativistic effects which vary in time is

$$\Delta t_\mathrm{r} = +4.4428 \times 10^{-10}\,\mathrm{s}\,\sqrt{\mathrm{m}}^{-1} e\sqrt{a}\sin E\,. \tag{24.43}$$

This correction of (24.43) is called the *eccentricity correction*; it is of the same form for all orbiting clocks and is ordinarily made by the receiver software. It represents 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 for GALILEO is about 24 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. For the QZS-1 satellite, the amplitude of this effect is about 200 ns. Figure 24.2 gives a plot of the relativistic effect – the negative of the correction.

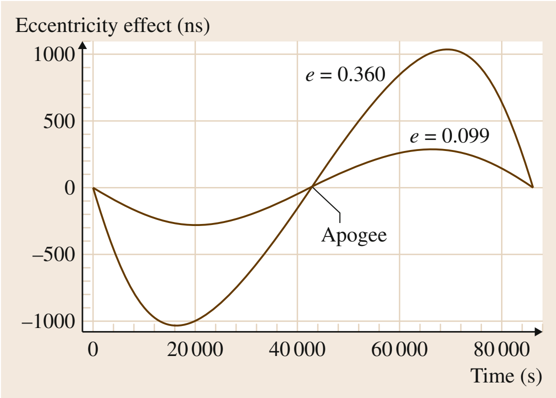

Fig. 24.2 Relativistic correction for orbital eccentricity effect, for a semimajor axis of 26 600 km

Equation (24.43) can be expressed without approximation in the following form, which is valid for Keplerian orbits,

$$\Delta t_r = +\frac{2\boldsymbol{r}\cdot\boldsymbol{v}}{c^2}\,, \quad (24.44)$$

where $\boldsymbol{r}$ and $\boldsymbol{v}$ are the position and velocity of the satellite at the instant of transmission. This may be proved using the expressions (24.33)–(24.36) for the Keplerian orbits of the satellites. This latter form is usually used in implementations of the receiver software.

It is not 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 [24.20].

24.7 Doppler Effect

Since orbiting clocks have had their rate adjusted so 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 $V(\boldsymbol{r}_j) \equiv V_j$, and $\boldsymbol{v}_j$, respectively. Let the frequency of the satellite transmission, before the rate adjustment is done, be f_0. After taking into account the rate adjustment discussed previously, it is straightforward to show that for a receiver of velocity $\boldsymbol{v}_R$ and gravitational potential V_R (in ECI coordinates), the received frequency f_R is given by

$$\begin{aligned}\frac{f_R - f_0}{f_0} &= \left[\frac{-V_R + \boldsymbol{v}_R^2/2 + \Phi_0 + 2GM_E/a + 2V_j}{c^2}\right] \\ &\quad \times \frac{(1 - \boldsymbol{N}\cdot\boldsymbol{v}_R/c)}{(1 - \boldsymbol{N}\cdot\boldsymbol{v}_j/c)}\,, \end{aligned} \quad (24.45)$$

where $\boldsymbol{N}$ 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

$$\frac{f_R - f_0}{f_0} = \left[\frac{2GM_E}{c^2}\left(\frac{1}{a} - \frac{1}{r}\right)\right]\frac{(1 - \boldsymbol{N}\cdot\boldsymbol{v}_R/c)}{(1 - \boldsymbol{N}\cdot\boldsymbol{v}_j/c)}\,. \quad (24.46)$$

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 $e \times 10^{-10}$.

Part D | 24.8

24.8 Relativity and Orbit Adjustments

To deal with satellite failures, it is common to have spares parked out of the way in orbits close to the nominal satellite orbits of the system. Performance of the clocks in these spares are monitored but not broadcast to the general user. As these spare satellites are raised or lowered in altitude to place them in assigned slots or take them out of service, their clocks suffer relativistic frequency changes from a combination of velocity changes and gravitational frequency shifts. If the initial and final orbits can be described as Keplerian orbits, (24.38) gives for the fractional frequency effect (the negative of the correction)

$$\frac{f-f_0}{f_0} = -\frac{3GM_E}{2c^2 a} - \Phi_0 \,. \tag{24.47}$$

The defined potential on the geoid, Φ_0, does not depend on satellite position. If the semimajor axis changes by a small amount δa, there will be a change in the frequency that can be adequately described by differentiating (24.47)

$$\delta\left(\frac{f-f_0}{f_0}\right) = +\frac{3GM_E}{2c^2 a^2}\delta a \,. \tag{24.48}$$

This simple equation has been very successful in predicting frequency shifts due to small changes in the semimajor axis. For a discussion of several measurements of such shifts, see [24.20]. The magnitudes of frequency shifts induced by such orbit changes are typically a few parts in 10^{13}.

The factor $3/2$ in (24.48) arises from the combined effect of second-order Doppler and gravitational frequency shifts. If the semimajor axis increases, the satellite will be higher in earth's gravitational potential and will be gravitationally blueshifted 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 redshift). The net effect would make a positive contribution to the fractional frequency shift.

24.9 Effects of Earth's Quadrupole Moment

Perturbations of GNSS orbits due to earth's quadrupole mass distribution are a significant fraction of the change in the semimajor 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 semimajor 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. Previously, such an effect on the SV clocks was neglected, and indeed it does turn out to be small. However, the effect is of the same order as the stability of the best orbiting clocks, so it is significant.

To see how large such quadrupole effects may be, we use exact calculations available in the literature, for the perturbations of the Keplerian orbital elements [24.16]. For the semimajor axis, if the eccentricity is very small the dominant contribution has a period twice the orbital period and has amplitude $3J_2a_1^2\sin^2 i/(2a)$, where a_1 is earth's equatorial radius and i is the inclination of the satellite orbit. The amplitude can be more than a kilometer.

The oscillation in the semimajor axis would significantly affect calculations of the radius at any particular time. This suggests that (24.37) needs to be reexamined in light of the periodic perturbations on the semimajor axis. Therefore, in this section we develop an approximate description of a satellite orbit, for 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 nontrivial 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.

24.9.1 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

$$V(x,y,z) = -\frac{GM_E}{r}\left(1-\frac{J_2a_1^2}{r^2}\left[\frac{3z^2}{2r^2}-\frac{1}{2}\right]\right). \tag{24.49}$$

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

$$\begin{aligned}\epsilon &= \text{constant}\\ &= \frac{v^2}{2} + V(x,y,z)\\ &= \frac{v^2}{2} - \frac{GM_E}{r} + V'(x,y,z)\,,\end{aligned} \tag{24.50}$$

where $V'(x,y,z)$ is the perturbing potential due to the earth's quadrupole potential.

It is shown in textbooks [24.16] that, with the help of Lagrange's planetary perturbation theory, the conservation of energy condition can be put in the form

$$\epsilon = -\frac{GM_E}{2a} + V'(x,y,z)\,, \tag{24.51}$$

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

$$\frac{v^2}{2} - \frac{GM_E}{r} = -\frac{GM_E}{2a}\,. \tag{24.52}$$

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 redshift) and a gravitational redshift. The total relativistic fractional frequency shift (relative to a reference at infinity) is

$$\frac{\Delta f}{f} = -\frac{v^2}{2} - \frac{GM_E}{r} + V'(x,y,z)\,. \tag{24.53}$$

The conservation of energy condition can be used to express the second-order Doppler shift in terms of the potential. 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 [24.16] 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.

24.9.2 Perturbed Semimajor Axis

From [24.16], the perturbed semimajor axis in the limit of negligible eccentricity is

$$a = a_m + \frac{3J_2a_1^2}{2a_m}\sin^2 i\cos(2nt+2\omega)\,, \tag{24.54}$$

where $n = \sqrt{GM_E/a_m^3}$ is the unperturbed mean motion, a_m is the mean semimajor axis, i the mean inclination, $n = \sqrt{GM_E/a_m^3}$ the unperturbed mean motion, and ω the mean altitude of perigee.

24.9.3 Perturbed Radius

The orbit radius depends on the combination $e\cos E$ where E is the eccentric anomaly. The eccentric anomaly depends on the mean anomaly; perturbation equations for the mean anomaly have terms with a factor e^{-1}, so one must take extra care in computing the product $e\cos E$ in order to obtain a meaningful result in the limit of small eccentricity. For the perturbed radius we then obtain

$$\begin{aligned}r &= a_m(1-e_m\cos E_m)\\ &\quad - \frac{3J_2a_1^2}{2a_m}\sin^2 i\cos(2nt+2\omega)\,.\end{aligned} \tag{24.55}$$

24.9.4 Perturbed Velocity

Then conservation of energy, (24.50) gives the following expression for the velocity

$$\begin{aligned}\frac{v^2}{2} &= \frac{GM_E(1+e_m\cos E_m)}{2a_m(1-e_m\cos E_m)}\\ &\quad + \frac{3GM_EJ_2a_1^2}{2a_m^3}\left(1-\frac{3}{2}\sin^2 i\right)\\ &\quad + \frac{GM_EJ_2a_1^2}{2a_m^2}\sin^2 i\cos(2nt+2\omega)\,.\end{aligned} \tag{24.56}$$

24.9.5 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 in $V'(x,y,z)$ which yields the expression

$$\begin{aligned}V'(x,y,z) &= -\frac{GM_EJ_2a_1^2}{2a_m^3}\left(1-\frac{3}{2}\sin^2 i\right)\\ &\quad - \frac{3GM_EJ_2a_1^2\sin^2 i}{4a_m^3}\cos(2nt+2\omega)\,.\end{aligned} \tag{24.57}$$

24.9.6 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

$$\begin{aligned}\frac{\Delta f}{f} &= -\frac{v^2}{2c^2} - \frac{GM_{\mathrm{E}}}{c^2 r} + \frac{V'}{c^2} \\ &= -\frac{3GM_{\mathrm{E}}}{2a_m c^2} - \frac{2GM_{\mathrm{E}}}{c^2 a_m}\frac{e_m \cos E_m}{1 - e_m \cos E_m} \\ &\quad - \frac{7GM_e J_2 a_1^2}{2a_m^3 c^2}\left(1 - \frac{3}{2}\sin^2 i\right) \\ &\quad - \frac{GM_{\mathrm{E}} J_2 a_1^2 \sin^2 i}{a_m^3 c^2}\cos(2nt + 2\omega)\,.\end{aligned} \tag{24.58}$$

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 often be neglected. The angle of inclination for which the third term vanishes exactly is $i = 55°$. For good coverage in the temperate zones, the orbits of most satellite navigation systems have inclinations very close to this value. For GPS the last term has an amplitude

$$\frac{GM_{\mathrm{E}} J_2 a_1^2 \sin^2 i}{a_m^3 c^2} = 6.95 \times 10^{-15}\,. \tag{24.59}$$

The best clocks in orbit in the GPS have stabilities of around 5 parts in 10^{15} at 1 day; this is only slightly less than the quadrupole effect, suggesting that this deterministic effect should be included in the systematic error budget.

The last periodic term in (24.58) is of a form similar to that which gives rise to the eccentricity correction, which is applied by GNSS receivers. Considering only the last periodic term, the additional time elapsed on the orbiting clock will be given by

$$\begin{aligned}\delta t_{J_2} = \int\limits_{\text{path}} \mathrm{d}t \Bigg[&-\frac{GM_{\mathrm{E}} J_2 a_1^2 \sin^2 i}{a_m^3 c^2} \\ &\times \cos(2nt + 2\omega)\Bigg]\,.\end{aligned} \tag{24.60}$$

Upon 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

$$\begin{aligned}\delta t_{J_2} = &-\sqrt{\frac{GM_{\mathrm{E}}}{a_m^3}}\frac{J_2 a_1^2 \sin^2 i}{2c^2} \\ &\times \sin(2nt + 2\omega)\,.\end{aligned} \tag{24.61}$$

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

$$\begin{aligned}\delta t_{J_2}(\text{correction}) = &\sqrt{\frac{GM_{\mathrm{E}}}{a_m^3}}\frac{J_2 a_1^2 \sin^2 i}{2c^2} \\ &\times \sin(2nt + 2\omega)\,.\end{aligned} \tag{24.62}$$

The phase of this correction is zero when the satellite passes through earth's equatorial plane going northwards.

24.9.7 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 [24.3]. The gravitational frequency shift of most GNSS satellite clocks from such potentials is a few parts in 10^{16}. However, this potential causes orbit perturbations of GNSS satellites that change both the radius in the main potential term $-GM_{\odot}/r$ and in the velocity; thus there are three contributions to the net frequency shift arising from this tidal potential. The geometry is complicated because earth's equatorial plane, the satellite orbital plane, and the ecliptic are inclined with respect to each other. Furthermore, there is a similar set of contributions from the moon's tidal potential that is larger and that can add to or subtract from solar tidal effects in a time-dependent manner. The net fractional frequency shift on a GALILEO satellite is estimated to be about five parts in 10^{15}.

24.10 Secondary Relativistic Effects

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

24.10.1 Signal Propagation Delay

The Shapiro signal propagation delay may be easily derived in the standard way from the metric, (24.25), which incorporates the choice of coordinate time rate expressed by the presence of the term in Φ_0/c^2. Setting $\mathrm{d}s^2 = 0$ and solving for the increment of coordinate time along the path increment

$$\mathrm{d}\sigma = \sqrt{\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2}$$

gives

$$\mathrm{d}t = \frac{1}{c}\left[1 - \frac{2V}{c^2} + \frac{\Phi_0}{c^2}\right]\mathrm{d}\sigma\ . \tag{24.63}$$

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

$$\Delta t_{\text{delay}} = \frac{\Phi_0}{c^2}\frac{l}{c} + \frac{2GM_\mathrm{E}}{c^3}\ln\left[\frac{r_1 + r_2 + l}{r_1 + r_2 - l}\right], \tag{24.64}$$

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 GNSS 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.

24.10.2 Effect on Geodetic Distance

At the level of a few millimeters, spatial curvature effects should be considered. For example, using the metric (24.26), the proper distance between a point at radius r_1 and another point at radius r_2 directly above the first is approximately

$$\int_{r_1}^{r_2}\mathrm{d}r\left[1 + \frac{GM_\mathrm{E}}{c^2 r}\right] = r_2 - r_1 + \frac{GM_\mathrm{E}}{c^2}\ln\left(\frac{r_2}{r_1}\right). \tag{24.65}$$

Between earth's surface and the radius of a geosynchronous satellite, the difference between proper distance and coordinate distance, and between the earth's surface and the radius of GPS satellites, is approximately 8 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.

24.10.3 Phase Wrap-Up

Transmitted signals from GNSS 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 the literature this is described in terms of an accumulation of phase called *phase wrap-up*. This effect has been experimentally measured with receivers spinning at rotational rates as low as 8 Hz [24.21, 22]. It is similar to an additional Doppler effect; it does not affect navigation if four signals are received simultaneously by the receiver as in (24.1).

24.11 Conclusions

GNSS is a remarkable laboratory for applications of the concepts of special and general relativity. It is also valuable as an outstanding source of pedagogical examples. It is particularly important to confirm that the basis for synchronization is on a firm conceptual foundation.

Plans are being made to put a laser-cooled clock having stability of $5\times10^{-14}/\sqrt{\tau}$ and accuracy of 1×10^{16}, on the international space station [24.23]. This will open up additional possibilities for testing relativity as well as for making improvements in GNSS.

References

24.1 J.J. Spilker Jr., B.W. Parkinson: Overview of GPS operation and design. In: *Global Positioning System: Theory and Applications*, (American Institute of Aeronautics and Astronautics, Washington 1996) p. 33

24.2 S. Schlamminger, K.-Y. Choi, T.A. Wagner, J.H. Gundlach, E.G. Adelberger: Test of the equivalence principle using a rotating torsion balance, Phys. Rev. Lett. **100**, 041101 (2008)

24.3 N. Ashby, B. Bertotti: Relativistic effects in local inertial frames, Phys. Rev. D **34**, 2246–2258 (1986)

24.4 R.A. Nelson: J., Math Phys. **28**, 2379–2383 (1987)

24.5 R.A. Nelson: J., Math Phys. **35**, 6224–6225 (1994)

24.6 S. Malys, J. Slater: Maintenance and enhancement of the world geodetic system 1984, Proc. ION-GPS-94 (1994) pp. 17–24

24.7 National Imagery and Mapping Agency Technical Report 8350.2, World Geodetic System 1984, Third Edition, Amendment 1, NIMA Stock No. DMATR83502WGS84, NSN 7643-01-402-0347

24.8 Addendum to NIMA TR 8350.2: Implementation of the World Geodetic System 1984 (WGS 84), Reference Frame G1150 (National Geospatial-Intelligence Agency 2001)

24.9 N. Ashby, M. Wess: *Global Positioning Receivers and Relativity*, NIST Technical Note, Vol. 1385 (U.S. Government Printing Office, Washington 1999)

24.10 N. Ashby: *An Earth-Based Coordinate Clock Network*, NBS Technical Note, Vol. 659 (U.S. Dept. of Commerce, U. S. Government Printing Office, Washington 1975), 20402 (S. D. Catalog # C13:46:659)

24.11 G. Kaplan: The IAU Resolutions on astronomical reference systems, time scales, and earth rotation models, US Naval Observatory Circular **179** (2005)

24.12 K. Lambeck: *Geophysical Geodesy*, Oxford Science Publications (Clarendon, Oxford 1988) pp. 13–18

24.13 G.D. Garland: *The Earth's Shape and Gravity* (Pergamon, New York 1965)

24.14 N. Ashby, J.J. Spilker Jr.: Introduction to relativistic effects on the Global Positioning System. In: *Global Positioning System: Theory and Applications*, (American Institute of Aeronautics and Astronautics, Washington 1996) pp. 623–697

24.15 N. Ashby, D.W. Allan: Practical implications of relativity for a global coordinate time scale, Radio Sci. **14**, 649–669 (1979)

24.16 P. Fitzpatrick: *The Principles of Celestial Mechanics* (Academic, New York 1970)

24.17 NAVSTAR GPS Space Segment/Navigation User Interfaces, ICD-GPS-200, Revision C (ARINC Research Corporation, Fountain Valley 1993)

24.18 GLONASS Interface Control Document, Edition 5.1 (Moscow 2008) p. 14

24.19 GALILEO Open Service Signal in Space Interface Control Document, Issue 1.1 (2010)

24.20 Ashby N.: Relativistic Effects in the global positioning system, available online at http://www.relativitylivingreviews.org/Articles/lrr-2003-1 (2007)

24.21 J.D. Kraus: *Antennas*, 2nd edn. (McGraw-Hill, New York 1988), reprinted by Cygnus-Quasar Books, Powell, Ohio

24.22 A.K. Tetewsky, F.E. Mullen: Effects of platform rotation on GPS with implications for GPS simulators, Proc. ION-GPS-96 (1996) pp. 1917–1925

24.23 C. Salomon, L. Cacciapuoti, N. Dimarcq: atomic clock ensemble in space: Fundamental physics and applications, Int. Jour. Mod. Phys. D **16**, 2511 (2007)