Relativity of GPS measurement

Thomas B. Bahder*

U. S. Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland 20783-1197, USA (Received 16 June 2003; published 29 September 2003)

The relativity of Global Positioning System (GPS) pseudorange measurements is explored within the geometrical optics approximation in the curved space-time near Earth. A space-time grid for navigation is created by the discontinuities introduced in the electromagnetic field amplitude by the *P*-code broadcast by the GPS satellites. We compute the world function of space-time near Earth, and we use it to define a scalar phase function that describes the space-time grid. We use this scalar phase function to define the measured pseudorange, which turns out to be a two-point space-time scalar under generalized coordinate transformations. Though the measured pseudorange is an invariant, it depends on the world lines of the receiver and satellite. While two colocated receivers measure two different pseudoranges to the same satellite, they obtain the correct position and time, independent of their velocity. We relate the measured pseudorange to the geometry of space-time and find corrections to the conventional model of the pseudorange that are on the order of the gravitational radius of the Earth.

DOI: 10.1103/PhysRevD.68.063005

PACS number(s): 95.30.Sf, 91.10.By, 95.55.Br

I. INTRODUCTION

Clock synchronization algorithms play a key role in applications such as communication, message encryption, and navigation. Recently, there have been a number of proposed clock synchronization algorithms based on a quantum information approach [1-8]. These discussions have been based mostly on nonrelativistic quantum mechanics. In many applications, however, clock synchronization must be performed between two nodes that are in relative motion, such as a jet aircraft and the ground, or between two nodes that are at different gravitational potentials, such as a satellite and the ground, or even two satellites at different altitudes. In such cases, a correct treatment of quantum clock synchronization must include relativistic effects from the start. The well-known tension between quantum mechanics and relativity theory makes this a difficult task. It is clear that the concepts of measurement, and transformation of measurable quantities under Lorentz transformations, play a key role in the problem of clock synchronization for both quantum and classical schemes. However, the transformation of measurable quantities has not been discussed even for the case of classical clock synchronization. In this article, we analyze in some detail the relativity of clock synchronization in the Global Positioning System (GPS), which is based on a classical synchronization scheme. In particular, we describe the transformation properties of the measured quantity called the pseudorange in the GPS [9-11]. We hope that clarifying the transformation properties of measurable quantities in classical clock synchronization will provide some useful insight into the problem of quantum clock synchronization.

The Global Positioning System (GPS) is a U.S. military constellation of satellites used for time keeping, and for navigation of land, air and sea [9-12]. Recently, two papers have analyzed the system of space-time coordinates that is used in the GPS [13,14]. In this paper, we address a different but

related aspect: the transformation of GPS pseudorange measurements. A remarkable aspect of the GPS is that a receiver need not be stationary with respect to the Earth's surface to obtain accurate time and position. A ship, a jet aircraft, or a low-Earth-orbit satellite can each compute accurate time and position, even though they have different velocities. This feature of the GPS is a consequence of two aspects: the signal structure of the satellite broadcasts and the special type of measurement that a GPS receiver makes. The GPS satellite signals set up an invariant grid of 3-dimensional space-time hypersurfaces (light cones). Each hypersurface is uniquely marked by the satellite that generated the hypersurface and by the space-time coordinates of the event of generation of the light cone [13,14]. By measuring the pseudorange to four satellites, the GPS receiver essentially determines its position by identifying the four hypersurfaces that it intersects. The pseudorange measurement is independent of receiver motion, up to an additive constant. For this reason, the measured pseudorange may be called a Lorentz pseudo-invariant. Below, we will see that the pseudorange is actually a two-point scalar under generalized coordinate transformations. The transformation properties of the pseudorange is a key element of the GPS, yet this subject has only briefly been mentioned in the literature [15].

In order to present a coherent description of the transformation properties of the pseudorange, we must deal with the nature of the broadcast GPS signals, the relativistic effects that impact these signals, as well as the measurement process itself. Therefore, the outline of this article is as follows. In Sec. II we present a two-receiver thought experiment to clarify the concept of invariance of pseudorange measurements. Section III contains a description of the space-time metric in the vicinity of the Earth. Section IV describes the relativistic effects on the GPS satellite clocks and on the observed signals, using the metric in Sec. III. Section V discusses the nature of the GPS broadcast signals, which are used to set up the space-time grid. Section VI discusses the pseudorange measurement process, using a mechanical analogue for a GPS receiver. The transformation properties of

^{*}Email address: bahder@arl.army.mil



FIG. 1. The world lines of two receivers, O_1 and O_2 , are shown. The receivers are tracking the same four GPS satellites. At event M the receivers coincide in space-time but have different velocities.

the pseudorange are obtained in Sec. VI. Section VII contains a brief discussion of navigation and time transfer using GPS signals. A correction is derived to the conventional flat space-time model of pseudorange. Section IX contains a summary and comments.

II. TWO RECEIVER EXPERIMENT

In order to make clear the concept of transformation of pseudorange measurements we offer a simple thought experiment. Consider two identical GPS receivers that are in relative motion. For example, one receiver is stationary with respect to the Earth's surface and the other is on a jet aircraft travelling at 1000 km/h. Each GPS receiver carries an identical copy of the software that is used to compute receiver position and time. Assume that each receiver is tracking the same four GPS satellites. See Fig. 1. Assume the two receivers' world lines cross at an event M in space-time. Does each receiver compute the same spatial position and time for the coincident event M? From a physics standpoint, the GPS satellites orbit the Earth at approximately 8.37 km/s, and there are large Doppler frequency shifts due to satellite and receiver motion [16]. In the GPS, the actual computation of receiver position and time depends on the space-time coordinates of the signal emission event (at the satellite), x_s^i , and reception event (at the receiver), x_o^i . In an Earth-centered inertial (ECI) frame, these coordinates are x_s^i and x_o^i ; however, in the receiver's comoving frame the emission and reception events have different coordinates, say x_{o}^{i} and $x_{o}^{\prime i}$. The two sets of coordinates, (x_{s}^{i}, x_{o}^{i}) and $(x_{s}^{\prime i}, x_{o}^{\prime i})$, for *i* =0,1,2,3, are related by a Lorentz transformation, which depends on receiver velocity with respect to the ECI frame. The actual electromagnetic field is different in each receiver's comoving frame, so there is different input information into each identical measuring device (receiver computer program). How can the two receivers compute the same spatial position and time from different input information? The key concepts are the space-time grid that is created by GPS satellites and the transformation properties of the measurable quantity in the GPS, which is called the pseudorange. These two themes are developed in the following sections.

III. CHOICE OF METRIC

In order to discuss relativistic effects, a reference frame or system of coordinates must be chosen. In the weak field limit, the metric of space-time in the vicinity of the Earth is of the form [17]

$$-ds^{2} = g_{ij}dx^{i}dx^{j} = -\left(1 + \frac{2}{c^{2}}V\right)(d\bar{x}^{0})^{2} + \left(1 - \frac{2}{c^{2}}V\right) \times [(dx^{1})^{2} + (dx^{2})^{2} + (dx^{3})^{2}], \quad (1)$$

where *V* is the Newtonian gravitational potential and $(\bar{x}^0, x^1, x^2, x^3)$ are the coordinates. The frame of reference in Eq. (1) can be thought of as an Earth-centered inertial (ECI) frame. We neglect small off-diagonal terms $g_{0\alpha}$, $\alpha = 1,2,3$, due to the rotation of the Earth.

In general relativity, the coordinates are mathematical entities that are never directly observed. However, it is useful to choose the coordinates in some physically meaningful way. The coordinates \bar{x}^0, x^1, x^2, x^3 are geocentric coordinates, where x^3 coincides with the Earth's axis of rotation and increasing positive values point to North. The Earth is modelled as an oblate spheroid with Newtonian potential given by [18]

$$V(r,\theta) = -\frac{GM}{r} \left[1 - J_2 \left(\frac{R}{r}\right)^2 P_2[\cos\left(\theta\right)] \right], \qquad (2)$$

where $r^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$ and θ is the polar angle measured from the x^3 axis. In Eq. (2), *G* is Newton's gravitational constant, *M* is the mass of the Earth, $P_2(x) = (3x^2 - 1)/2$ is the second Legendre polynomial, *R* is the Earth's equatorial radius, and J_2 is the Earth's quadrupole moment, whose value is approximately $J_2 = 1.08 \times 10^{-3}$, see Table I. The metric in Eq. (1) is the solution of the linearized Einstein field equations [17]. In Eq. (1), the coordinate time \bar{x}^0 has no simple relation to the time kept by ideal clocks on the surface of the Earth.

The coordinate time can be given a simple interpretation by transforming the metric in Eq. (1) to rotating Earthcentered Earth-fixed (ECEF) coordinates y^i , using the transformation

$$\overline{x}^{0} = \overline{y}^{0}$$

$$x^{1} = \cos\left(\frac{\omega}{c}\overline{y}^{0}\right)y^{1} - \sin\left(\frac{\omega}{c}\overline{y}^{0}\right)y^{2}$$

$$x^{2} = \sin\left(\frac{\omega}{c}\overline{y}^{0}\right)y^{1} + \cos\left(\frac{\omega}{c}\overline{y}^{0}\right)y^{2}$$

$$x^{3} = y^{3}.$$
(3)

Symbol	Definition	Value	Units
GM	gravitational constant times Earth's mass	3.986004415×10 ¹⁴ (JGM-2)	m^3/s^3
с	vacuum speed of light	2.99792458×10^8 (exact definition)	m/s
Ω	angular velocity of Earth rotation	$7.2921151467 \times 10^{-5}$ (WGS-84)	radian/s
R	mean Earth radius at the Equator	6.3781363×10 ⁶ (JGM-2)	m
J_2	Earth's quadrupole moment	1.0826269×10^{-3} (JGM-2)	1
а	GPS orbit semimajor axis	26561.75×10 ³ (Ref. [19])	m
$v_s/c = \sqrt{GM/a} / c$	GPS satellite velocity/ c	1.29217×10^{-5}	1
GM/Rc^2	gravitational potential scale	6.95348×10^{-10}	1

TABLE I. The numerical constants and values used in the calculations. Values taken from Refs. [9] and [20].

Note that the coordinate time in the rotating frame, \overline{y}^0 , is equal to the coordinate time in the ECI frame \overline{x}^0 . In these ECEF rotating coordinates, the metric is given by

$$-ds^{2} = -\left[1 + \frac{2V}{c^{2}} - \frac{\Omega^{2}}{c^{2}} [(y^{1})^{2} + (y^{2})^{2}] + \frac{2V}{c^{2}} \frac{\Omega^{2}}{c^{2}} [(y^{1})^{2} + (y^{2})^{2}]\right] (d\bar{y}^{0})^{2} + \left(1 - \frac{2V}{c^{2}}\right) \left[2\frac{\Omega}{c} (y^{1}dy^{2} - y^{2}dy^{1}) + (d\bar{y}^{0})^{2} + (dy^{1})^{2} + (dy^{2})^{2} + (dy^{3})^{2}\right].$$
(4)

From Eq. (4), we see that there exist geopotential surfaces, $\phi(y^1, y^2, y^3) = c$, where c is a constant,

$$\phi(y^1, y^2, y^3) = V - \frac{1}{2}\Omega^2[(y^1)^2 + (y^2)^2]$$
(5)

where stationary clocks in the ECEF frame (that satisfy $dy^{\alpha}=0$) have the same rate of proper time $d\tau=ds/c$ with respect to coordinate time \overline{y}^0 [19]. In other words, ideal clocks located at the same value of geopotential ϕ have the same rate with respect to coordinate time y^0 . We have neglected the small cross-term $2V\Omega^2R^2/c^4 \sim 10^{-21}$.

Using the observation that clocks at a constant value of geopotential run at the same rate, it is advantageous to define [19] the new coordinate time t

$$ct = x^{0} = y^{0} = \left(1 + \frac{\phi_{o}}{c^{2}}\right) \overline{y}^{0} = \left(1 + \frac{\phi_{o}}{c^{2}}\right) \overline{x}^{0}$$
(6)

where ϕ_o is the value of the geopotential ϕ on the Earth's equator, at $\theta = \pi/2$ and r = R:

$$\phi_o = -\frac{GM}{R} \left(1 + \frac{1}{2} J_2 \right) - \frac{1}{2} \Omega^2 R^2.$$
 (7)

For the values of the constants in Table I, the dimensionless magnitude of this term is $\phi_o/c^2 = -6.96928 \times 10^{-10}$. Using the transformation in Eq. (6), the metric in Eq. (4) becomes

$$-ds^{2} = -\left[1 + \frac{2}{c^{2}}(\phi - \phi_{o})\right](dy^{0})^{2} + \left(1 - \frac{2V + \phi_{o}}{c^{2}}\right)$$
$$\times 2\frac{\Omega}{c}(y^{1}dy^{2} - y^{2}dy^{1})dy^{0} + \left(1 - \frac{2V}{c^{2}}\right)[(dy^{1})^{2} + (dy^{2})^{2} + (dy^{3})^{2}]. \tag{8}$$

Equation (8) gives the space-time metric in ECEF rotating coordinates y^i . Note that an ideal clock that is stationary in ECEF coordinates (with $dy^{\alpha}=0$), has proper time

$$d\tau = ds/c = \frac{1}{c} \left[1 + \frac{2}{c^2} (\phi - \phi_o) \right]^{1/2} dy^0.$$
 (9)

When this clock is located on the geoid, then $\phi - \phi_o = 0$, and $d\tau = dy^0/c$, so this ideal clock keeps coordinate time, $x^0 = y^0$. Hence a good hardware clock that is on the geoid, and stationary with respect to the rotating Earth, can be used as a reference clock to keep coordinate time. Note that by Eq. (6) the coordinate time in rotating ECEF coordinates is the same as coordinate time in ECI coordinates, therefore, the same clock keeps coordinate time in the ECI frame, x^0 , and coordinate time in the ECEF frame, y^0 .

Using the coordinate time transformation in Eq. (6), the metric in Eq. (1) becomes

$$-ds^{2} = g_{ij}dx^{i}dx^{j}$$

$$= -\left[1 + \frac{2}{c^{2}}(V - \phi_{o})\right](dx^{0})^{2} + \left(1 - \frac{2}{c^{2}}V\right)$$

$$\times [(dx^{1})^{2} + (dx^{2})^{2} + (dx^{3})^{2}].$$
(10)

Equation (10) gives the metric in ECI coordinates. The coordinate time that enters into the metric, x^0 , is the time kept by ideal clocks on the geoid. This result was the goal of the time transformation given in Eq. (6). Note, however, that in the ECI frame metric in Eq. (10), the proper time interval dson a stationary clock in ECI coordinates (with $dx^{\alpha}=0$), is not equal to the coordinate time interval dx^0 because in general $V \neq \phi_o$. The ECI coordinate metric, given in Eq. (10), is useful for computing the proper time $d\tau = ds/c$ elapsed onboard a satellite clock, in terms of elapsed coordinate time.

IV. RELATIVISTIC EFFECTS IN GPS

A. Satellite oscillator frequency offset

The clocks in GPS satellites are at a higher gravitational potential than the clocks on Earth. As observed on the Earth, this difference in the gravitational potential causes the oscillators of GPS atomic clocks to appear to run fast, by fractional frequency [17]

$$\frac{\Delta\omega}{\omega} = \frac{\phi_1 - \phi_2}{c^2} \tag{11}$$

where the approximate gravitational potential $\phi = -GM/r$, and *r* is the distance from the center of the Earth. Here, ϕ_1 and ϕ_2 are the potentials at the satellite and on the earth surface, respectively. For GPS, $\Delta \omega / \omega \approx 5.28 \times 10^{-10}$. This effect causes the satellites clocks to run fast by 45 μ s per day. This is often called a gravitational redshift, but actually, it is a blueshift (toward higher frequencies).

In addition to the gravitational frequency shift, the GPS satellites are moving. Consequently, as observed in the ECI frame, satellite oscillators exhibit time dilation due to their velocity $v/c \approx 8.37$ km/s. The time dilation effect in special relativity is given by

$$\Delta t = \gamma \Delta t' \tag{12}$$

where Δt is the time interval in the ECI frame and $\Delta t'$ is the proper time of a clock moving at speed v in the ECI frame. For GPS satellites, $\gamma \approx 1 - v^2/(2c^2) \approx 8.33 \times 10^{-11}$. With respect to coordinate time in the ECI frame, the time dilation effect makes the satellite clocks appear to run slow by approximately 37 μ s per day.

The typical GPS atomic clock stability is 1 part in 10^{-13} , so the effect of time dilation and gravitational redshift are each about 10^3 times larger, and therefore, both effects must must be taken into account. The net effect of time dilation and gravitational redshift is that the atomic clocks would run fast by 38 μ s per day (=45 μ s-7 μ s). This is a huge effect, which can be measured by the fact that 38 μ s corresponds to a range error of 38,000 feet per day.

The actual value of the combined effect of the gravitational potential and time dilation on the frequency of the satellite oscillator is computed using the metric in Eq. (10). During a coordinate time dx^0 , the satellite moves a spatial distance dx^{α} , $\alpha = 1,2,3$. The proper time elapsed on the satellite clock, $d\tau_s = ds/c$, is related to elapsed coordinate time, dx^0 , by

$$ds = c d\tau_s = \left[1 + \frac{2}{c^2}(V_s - \phi_o) - \left(1 - \frac{2V_s}{c^2}\right)v_s^2\right]^{1/2} dx^0$$
(13)

where V_s is the Earth's gravitational potential, given in Eq. (2), evaluated at the position of the satellite and $v_s^2 = \delta_{\alpha\beta}(dx^{\alpha}/dx^0)(dx^{\beta}/dx^0)$ is the square of the satellite velocity divided by c^2 . In Eq. (13), we take the Earth's quadrupole potential to be zero, $J_2=0$, which allows a circular orbit for the satellite and makes $d\tau_s/dx^0$ independent of the

polar angle θ of the satellite. Similarly, we approximate the satellite velocity v_s by taking $J_2=0$ and assume a circular (zero eccentricity) orbit so that $v_s^2 = GM/ac^2$. Expanding Eq. (13) to first order in small quantities, V_s/c^2 and ϕ_o/c^2 , we have

$$c\frac{d\tau_s}{dx^0} = 1 - \frac{3GM}{2c^2a} - \frac{\phi_{oo}}{c^2} = 1 + \delta$$
(14)

where

$$\phi_{oo} = -\frac{GM}{R} - \frac{1}{2}\Omega^2 R^2 \tag{15}$$

is the value of ϕ_o when $J_2=0$. Equation (14) gives an approximate expression for the rate of proper time on a GPS satellite in circular orbit, with respect to coordinate time x^0 in the metric in Eq. (10). We have dropped terms $V_s v_s^2/c^2 \approx 10^{-20}$ in Eq. (14), since $v_s^2 \approx 10^{-10}$ and $V_s/c^2 \approx 10^{-10}$.

The ratio of the frequencies of oscillators is inversely proportional to the clock rates, $\omega_{coord}/\omega_s = d\tau_s/dx^0$, where ω_{coord} is the number of cycles elapsed as counted per unit of coordinate time dx^0 , and ω_s is the number of cycles elapsed as counted per unit of proper time $d\tau_s$ on-board the satellite. Therefore, the oscillators of clocks on-board the GPS satellites in circular orbit (taking $J_2=0$)have a frequency shift [19]

$$\frac{\Delta\omega}{\omega_s} = \frac{\omega_{coord} - \omega_s}{\omega_s} = -\frac{3GM}{2c^2a} - \frac{\phi_{oo}}{c^2} \equiv \delta_0.$$
(16)

Using the value of JGM-2 constants [20]

$$\delta_0 = +4.460963 \times 10^{-10}. \tag{17}$$

If the clocks on-board GPS satellites were allowed to run freely, then Eq. (14) shows that during one day of elapsed coordinate time Δx^0 , a satellite clock would gain $c(d\tau_s/dx^0-1) \times 24$ h. Nominally, the GPS system is designed to transmit the digital P-code at 10.23 MHz. However, Eq. (16) shows that if the oscillator in the satellite were set to $\omega_s/2\pi = 10.23$ MHz, then this code would appear to have a higher frequency, $\omega_{coord} = (1 + \delta)\omega_s$, as measured with respect to a clock that keeps coordinate time in the ECI frame. If a clock is stationary in the ECEF frame, and is located on the Earth's geoid, it keeps coordinate time x^0 , and therefore the satellite clocks would appear to run fast to a GPS user on the geoid. Consequently, in the GPS, the technical specifications for the satellite clocks include a frequency "factory offset" that is applied prior to the launch of the satellites. The actual (angular) frequency of the P-code that is broadcast by the satellite clock is [21]

$$\omega_s = 2\pi (1 - \delta_0) f_0 \tag{18}$$

where $\delta_0 = +4.46 \times 10^{-10}$ and $f_0 = 10.23$ MHz are fixed GPS constants. With this frequency correction applied, the satellite clocks approximately keep coordinate time x^0 in the ECI frame.

However, satellite orbits cannot be made perfectly circular, so the GPS clocks are in a slightly eccentric orbit. This eccentricity of the orbit contributes an additional well-known effect on the time of the satellite clocks: the satellite clocks periodically speed up and slow down with respect to coordinate time x^0 . This effect is often called the "e Sin E effect" and depends on the position of the satellite in its orbit. This clock correction is made in the GPS receiver in software, and is given by [22]

$$\Delta t_r = \frac{2}{c^2} \sqrt{GMa} \ e \sin E \tag{19}$$

where *a* is the semimajor axis of the satellite, *e* is the orbital eccentricity, and *E* is the eccentric anomaly along the orbit. For a typical upper limit of e = 0.01, $\Delta t_r \approx 23$ ns.

B. Observed frequency shift

Equation (18) gives the frequency offset that is applied to a GPS satellite clock/oscillator so that it approximately keeps coordinate time x^0 , with respect to a reference oscillator that is stationary in the ECEF frame and located on the geoid. Most users of the GPS do not satisfy these two criteria. For example, a user that is stationary in the ECEF frame on the geoid is moving in the ECI frame, and a user in an aircraft is above the geoid. Therefore, at any given time, a user typically sees a frequency shift that is different from the "factory offset," given in Eq. (18). The actual measured frequency shift of the satellite signal depends on user and satellite positions in the gravitational field of the Earth, and also on user and satellite velocities (not just user-satellite velocity differences). The reason that the frequency shift depends separately on user and satellite velocities is, of course, that spacetime is not homogeneous (space-time is not flat) because of the Earth's gravitational field.

Consider a satellite moving at velocity \mathbf{v}_s at gravitational potential V_s . At event *S*, the satellite transmits an electromagnetic signal of proper frequency ω_s , as measured with respect to a calibrated oscillator on-board the satellite. An observer at event *O* is at gravitational potential V_o and has a velocity \mathbf{v}_o . The observer measures a signal having frequency ω_o , which is different from ω_s because he is in motion and at a different gravitational potential than the satellite. The emission event *S* and observer reception event *O* are connected by a null geodesic. The quantity $\omega_o/\omega_s - 1$ is a 2-point scalar because it depends on two space-time points, *S* and *O*. A 2-point scalar transforms as a true scalar under separate transformations of coordinates at point *S* and at point *O* [23]. Using the metric in Eq. (10), a detailed calculation gives (see Appendix B)

$$\frac{\omega_o}{\omega_s} - 1 = (\mathbf{v}_s - \mathbf{v}_o) \cdot \mathbf{n} \left[1 + \frac{V_s - V_o}{c^2} + \frac{1}{2} (v_o^2 - v_s^2) + (\mathbf{v}_s \cdot \hat{\mathbf{n}}) \right] \\ + (\mathbf{v}_s \cdot \hat{\mathbf{n}})^2 - \alpha_1 - \beta_1 \left] + \frac{V_s - V_o}{c^2} + \frac{1}{2} (v_o^2 - v_s^2) \right] \\ + \frac{2GM}{c^2 r} \left[h(\mathbf{r}_o, \mathbf{r}_s, \mathbf{v}_o, \mathbf{v}_s) + h(\mathbf{r}_s, \mathbf{r}_o, \mathbf{v}_s, \mathbf{v}_o) \right]$$
(20)

where \mathbf{r}_o and \mathbf{r}_s are positions, and \mathbf{v}_o and \mathbf{v}_s are velocities (divided by *c*), of observer and satellite (source), respectively. In Eq. (20), $r = |\mathbf{r}_o - \mathbf{r}_s|$, $v_o^2 = \mathbf{v}_o \cdot \mathbf{v}_o$, $v_s^2 = \mathbf{v}_s \cdot \mathbf{v}_s$, and V_s and V_o are the Earth's gravitational potential given in Eq. (2), evaluated at the emission event *S* and reception event *O*, respectively. The unit vector $\hat{\mathbf{n}}$ connects the emission and reception events, *S* and *O*:

$$\hat{\mathbf{n}} = \frac{\mathbf{r}_o - \mathbf{r}_s}{|\mathbf{r}_o - \mathbf{r}_s|}.$$
(21)

The dimensionless quantities $\alpha_1 = GM/c^2R$ and $\beta_1 = \Omega^2 R^2/2c^2$. The last term in the right-hand side of Eq. (20) is given by the sum of two terms, which are related by interchanges of subscripts "s" and "o." The scalar function $h(\mathbf{r}_{\alpha}, \mathbf{r}_{s}, \mathbf{v}_{\alpha}, \mathbf{v}_{s})$ is given by

$$h(\mathbf{r}_{o}, \mathbf{r}_{s}, \mathbf{v}_{o}, \mathbf{v}_{s}) = \frac{1}{1 - (\hat{\mathbf{n}} \cdot \hat{\mathbf{r}}_{o})^{2}} \left\{ \left[\frac{r_{o}}{r} - \frac{\mathbf{r}_{o} \cdot \mathbf{r}_{s}}{rr_{o}} \right] \hat{\mathbf{n}} \cdot (\mathbf{v}_{o} - \mathbf{v}_{s}) - \frac{\mathbf{r}_{o}}{r_{o}} \cdot (\mathbf{v}_{o} - \mathbf{v}_{s}) + \frac{\mathbf{r}_{s} \cdot \mathbf{v}_{o}}{r_{o}} - \frac{(\mathbf{r}_{o} \cdot \mathbf{r}_{s})(\mathbf{r}_{o} \cdot \mathbf{v}_{o})}{r_{o}^{3}} \right\}.$$
(22)

For a GPS user on the surface of the Earth, $(V_s - V_o)/c^2 \approx 5.3 \times 10^{-10}$, and the GPS satellite velocity (divided by *c*) is $v_s \approx 1.3 \times 10^{-5}$. We also have that $\alpha_1 \approx 6.9 \times 10^{-10}$ and $\beta_1 \approx 1.2 \times 10^{-12}$. Therefore, in the derivation of Eq. (20), we have taken $v_s = O(1)$, $V_s/c^2 = O(2)$, $V_o/c^2 = O(2)$, $\alpha_1 = O(2)$, and $\beta_1 = O(2)$, where $O(1) \sim 10^{-5}$, and I have dropped terms $O(4) \sim 10^{-20}$.

The measured frequency is due to three types of terms. First, there is a special relativistic Doppler effect that depends on the relative velocity of the satellite and observer, contained in the terms proportional to $(\mathbf{v}_s - \mathbf{v}_o) \cdot \mathbf{n}$, and the term $\frac{1}{2}(v_o^2 - v_s^2)$, which comes from expanding the special relativistic γ factors for the satellite and observer. Next, there is a frequency shift due to the difference in gravitational potential of the observer and satellite, which is given by the stand alone term $(V_s - V_o)/c^2$. Finally, there are cross terms that depend on products of satellite and observer velocities and the Earth's mass M.

As described in Eq. (18), the GPS satellites have a built-in "factory offset," $\delta_0 = 4.46 \times 10^{-10}$. From Eq. (20) we see that the observed frequency shift of the satellite signal due to gravitational potential differences, $(V_s - V_o)/c^2$, is on the order of 10^{-10} and can vary depending on the altitude of the observer. In addition to this frequency shift, there is a Doppler frequency shift that is much larger. For example, for a jet aircraft travelling along the equator at approximately 1000 km/hour with respect to the Earth's surface, the velocity (fraction of c) with respect to the ECI coordinates can be on the order $v_o \sim 2 \times 10^{-6}$, which is a factor of 5×10^3 larger than the built-in factory offset. Similarly, for an observer on-board a low-Earth orbit satellite at altitude 1000 km, whose orbit is in the plane of the equator, we have $|\mathbf{v}_s - \mathbf{v}_o|$



FIG. 2. Each component of the electric field, $E(\tau^*)$, broadcast by a GPS satellite has possible discontinuities in its amplitude at the times, $\tau_1^*, \tau_2^*, \ldots, \tau_N^*$, where the pseudorandom P-code function $s(\tau^*)$ changes sign. These discontinuities propagate on 3D hypersurfaces in space-time and define the space-time grid.

 $\sim 10^{-5}$. This frequency shift is on the order of 10^4 times larger than the built-in factory offset. Of course the factory offset is applied to cancel out a secular effect—an effect that leads to a constantly increasing discrepancy in time between satellite and coordinate time clocks, while the Doppler (motional) effect has a more complicated time dependence. The point here is that the frequency of GPS signals, as seen by an observer, has huge frequency shifts due to observer motion. Below, we describe how these frequency shifts are essentially removed—so that the actual measurements made by a GPS receiver are independent of the velocity of the observer (GPS receiver).

V. GPS SIGNALS: THE SPACE-TIME GRID

The GPS satellites broadcast electromagnetic signals that set up a geometric space-time grid [24]. Users of the GPS that receive four satellite signals can uniquely identify their position in the space-time grid. This grid is created by discontinuities in the amplitude of the broadcast electromagnetic field.

All the GPS satellites broadcast on two carrier frequencies in the L-band centered about $L_1 \approx 1575.42$ MHz and $L_2 \approx 1227.6$ MHz. The GPS satellites have helical antennas that are constantly pointed toward the Earth's center. Near the antenna axis (bore site) the broadcast electromagnetic radiation field is right circularly polarized [15]. Superimposed on each satellite carrier signal is a unique code, or sequence of digital bits that identifies the satellite that is broadcasting. At a given time and spatial position in the comoving frame of the satellite, one vector component of the electric field can be written approximately as

$$E = D(\tau^*)s(\tau^*)\cos\left(q \ \omega_s \tau^* + \phi\right) \tag{23}$$

where ω_s is the P-code transmission frequency given in Eq. (18), τ^* is the hardware time kept on-board the satellite by its local clock, and ϕ is a phase associated with phase noise in the signal. The integer q takes values q = 154 or q = 120, to produce the broadcast signal that is transmitted on the two

carrier frequencies, $L_1 = 154 \times \omega_s/2\pi \approx 1575.42$ MHz or $L_2 = 120 \times \omega_s/2\pi \approx 1227.6$ MHz. See Fig. 2. The L_1 and L_2 carrier frequencies are integer multiples of the code bit transmission rate, $\omega_s/2\pi$. We distinguish between proper time τ kept by an ideal clock, and hardware time, τ^* , which is kept by a clock that is a real physical device. The function s(t) takes the discrete values -1 and +1, and represents the digital P-code values 0 and 1 in the pseudorandom code, which is unique to each satellite [21]. There are exactly $N (\approx 10^{12})$ values in the digital P-code sequence. The code sequence starts at midnight on Sunday and has a period of exactly one week: s(t+T)=s(t), where T is one week. Therefore, we can define a discrete phase Φ_n for the periodic function s(t) by

$$\Phi^{(n)} = 2\pi \frac{(n-1)}{N}$$
(24)

where n = 1, 2, 3, ..., N, is an integer that sequentially labels the bits in the code s(t).

In Eq. (23), the function $D(\tau^*)$ is a digital navigation message that is broadcast at approximately 50 Hz. The message provides the satellite ephemeris and satellite clock corrections in the form of two coefficients, A and B. The clock corrections essentially provide the conversion from hardware time τ^* to coordinate time in the ECI frame, x^0 , in the form $x^0 = A + B\tau^*$. The functions $s(\tau^*)$ and $D(\tau^*)$ are timed so that changes between 0 and 1 in $s(\tau^*)$ occur at the same point in time as in $D(\tau^*)$, i.e., the bit transitions (edges)in $D(\tau^*)$ align with those in $s(\tau^*)$. The function $D(\tau^*)$ provides the broadcast ephemeris of each satellite.

In the satellite comoving frame, the pseudorandom code, $s(\tau^*)$, is broadcast at the factory adjusted angular frequency ω_s given in Eq. (18), as timed by the atomic clock on-board each satellite. The frequency of the oscillator (clock) on-board the satellite has been lowered from f_0 MHz to $\omega_s/2\pi$ [see Eq. (18)] so that, as seen from the ECI frame of reference, the discontinuities in the amplitude of the electromagnetic field occur at time intervals of approximately $1/f_0$ with respect to coordinate time in the metric given in Eq. (10)

[25]. Since the code $s(\tau^*)$ is timed according to the satellite clock, the discontinuities in the code can occur at hardware times τ_n^* , n=1,2,3,...,N, where *n* sequentially labels the (potential) code edge discontinuities since midnight Sunday, and *N* is the number of code bits in the $s(\tau^*)$ code. All satellites have the same sequence of hardware times τ_n^* for possible code-edge-emission events, given by

$$\tau_n^* = \frac{\Phi^{(n)}}{\omega_s/N} = \frac{2\pi}{\omega_s}(n-1) \tag{25}$$

where ω_s is a constant frequency given in Eq. (18) and n $=1,2,3,\ldots,N$ is the bit number in the P-code sequence $s(\tau^*)$. In the ECI frame of reference, the coordinates of the world line of the antenna phase center for satellite s, $x_s^i(\tau^*)$, can be parametrized by the satellite clock's hardware time, τ^* . The world line of a satellite is approximately a geodesic; however, forces on a satellite, such as solar pressure, and the antenna phase center position differing from the satellite center of mass, lead to an effective world line that differs from a geodesic. The world line of the antenna phase center is precisely determined by ground tracking stations. Future points on the world line are computed and expressed in terms of classical satellite ephemeris parameters and the A and B clock corrections. This information is uploaded to the satellite's navigation message, which is transmitted to GPS users through the digital sequence $D(\tau^*)$.

It is well known that discontinuities in the emitted electromagnetic field tensor define invariant 3D hypersurfaces called characteristics [23,26,27]. In terms of the world function of the space-time, for each satellite, these hypersurfaces are given by [28]

$$\Omega(x_s^i(\tau_n^*), x^j) = 0, \quad n = 1, 2, 3, \dots, N$$
(26)

where $x_s^i(\tau^*)$ is the world line of the satellite *s* parametrized by satellite hardware time τ^* , and the 3D hypersurfaces are defined by their coordinates x^j , j=0,1,2,3, that satisfy Eq. (26) and are on the forward light cone, so that $x^0 > x_s^0(\tau_n^*)$. Each hypersurface can be uniquely specified by τ_n^* , the hardware time of satellite "s" and the bit number *n* associated with the (discontinuity) new bit in the code $s(\tau^*)$. See Fig. 3. For flat space-time with a Minkowski metric, Eq. (26) reduces to

$$\frac{1}{2} \eta_{ij} [x^i - x^i_s(\tau^*_n)] [x^j - x^j_s(\tau^*_n)] = 0$$
(27)

where η_{ij} is the Minkowski metric with nonzero diagonal components (-1,+1,+1,+1), and there is no sum on *n*.

A detailed calculation of the world function entering in Eq. (26) for the metric in Eq. (10) gives



FIG. 3. Discontinuities in the amplitude of the electromagnetic field at times, $\tau_1^*, \tau_2^*, \ldots, \tau_N^*$, propagate on 3D hypersurfaces in space-time that define the space-time grid.

$$\Omega(x_1^i, x_2^j) = -\frac{1}{2} (\Delta x^0)^2 \left[1 + \frac{2GM}{c^2 R} + \frac{\Omega^2 R^2}{c^2} - \frac{2GM}{c^2 |\Delta \mathbf{r}|} \Lambda(\mathbf{r}_1, \mathbf{r}_2) \right] + \frac{1}{2} (\Delta \mathbf{r})^2 \times \left[1 + \frac{2GM}{c^2 |\Delta \mathbf{r}|} \Lambda(\mathbf{r}_1, \mathbf{r}_2) \right]$$
(28)

where in Eq. (28) we use the following definitions: $x_1^i = (x_1^0, \mathbf{r}_1), x_2^i = (x_2^0, \mathbf{r}_2), \Delta x^0 = x_1^0 - x_2^0$ and $|\Delta \mathbf{r}| = |\mathbf{r}_1 - \mathbf{r}_2|$,

$$\Lambda(\mathbf{r}_1, \mathbf{r}_2) = \log \left(\frac{\tan\left(\frac{\theta_1}{2}\right)}{\tan\left(\frac{\theta_2}{2}\right)} \right)$$
(29)

and θ_1 and θ_2 are defined by

$$\cos \theta_a = \frac{\mathbf{r}_a \cdot (\mathbf{r}_2 - \mathbf{r}_1)}{|\mathbf{r}_a||\mathbf{r}_2 - \mathbf{r}_1|}, \qquad a = 1, 2.$$
(30)

In Eq. (28), we have taken the small parameter $J_2 = 0$.

Each GPS satellite broadcasts a set of 3D hypersurfaces that form part of the coordinate grid, given in Eq. (26). There are approximately 24 satellites in the GPS constellation, and all the hypersurfaces from these satellites comprise the GPS space-time grid that is used to label events in this space-time. The hypersurfaces are spaced approximately $1/f_0 \approx 97.75$ ns in coordinate time and 29.31 m in space; see the metric in Eq. (10). Since the coordinate hypersurfaces are so far apart in space and time, an event in this space-time can be specified more accurately by interpolating the times τ_n^* at which discontinuities in electromagnetic field are actually emitted. For each satellite, *s*, an interpolated hardware time, τ_{η}^* , and corresponding interpolated phase of pseudorandom code, Φ , can be associated with a real number η

$$\tau_{\eta}^{*} = \frac{\Phi}{\frac{\omega_{s}}{N}} = \frac{2\pi}{\omega_{s}} (\eta - 1)$$
(31)

where ω_s is given by Eq. (18), $\tau_n^* \leq \tau_\eta^* < \tau_{n+1}^*$ and $n \leq \eta < n+1$, see Eq. (25). The real number η is the interpolated value of the code bit integer number *n*. Associated with the interpolated hardware time, τ_η^* , there is a continuous family of coordinate 3D hypersurfaces analogous to those in Eq. (26):

$$\Omega(x_s^i(\tau_n^*), x^j) = 0. \tag{32}$$

The continuous parameter τ_{η}^* labels the 3D hypersurface that has code phase Φ and is defined by coordinates x^j that satisfy Eq. (32). On each 3D hypersurface, the phase Φ has the value

$$\Phi = \frac{\omega_s}{N} \tau_\eta^* \tag{33}$$

where ω_s is given by Eq. (18). The hardware time τ_{η}^* of an emission event at the satellite is related to coordinate time $t_s = x_s^0/c$ in the ECI frame of reference by the satellite clock correction $\Delta \tau_{\eta}^*$

$$t_s = \tau_\eta^* + \Delta \tau_\eta^* \,. \tag{34}$$

In terms of coordinate time of emission, the phase broadcast by satellite s is then given by

$$\Phi_s(t,\mathbf{r}) = \frac{\omega_s}{cN} [x_s^0(t,\mathbf{r}) - c\Delta\tau_\eta^*]$$
(35)

where $\omega_s \Delta \tau_{\eta}^*/N$ represents a phase offset due to the fact that satellite clocks keep hardware time, which is an approximation of coordinate time $t_s = x_s^0/c$. The broadcast phase function satisfies the eikonal equation [26]

$$g^{ij}\frac{\partial \Phi_s}{\partial x^i}\frac{\partial \Phi_s}{\partial x^j} = 0 \tag{36}$$

where g^{ij} are the contravariant components of the metric given in Eq. (10) (no sum on *s*). The phase function Φ_s depends on the invariant world line of the satellite, $x_s^i(\tau^*)$. From Eq. (36) it is clear that under Lorentz transformations, or generalized coordinate transformations, the phase $\Phi_s(t,\mathbf{r})$ transforms as a scalar [29]. The form of Eq. (36) shows that the wave vector associated with the phase Φ_s , κ_i $= \partial \Phi_s / \partial x^i$, is a null vector. The wave vector κ_i can be related to the covariant derivative of the world function of space-time, $\Omega(T,R)$, between emission event $T=(t_T,\mathbf{r}_T)$ and reception event $R=(t,\mathbf{r})$. The direction of the wave vector κ_i is the same as the direction of the covariant derivative of the world function, Ω_{i_s} , where the derivative is taken with respect to *R*. This can be seen from the identity $\Omega = g^{ij}(R)\Omega_{i_R}\Omega_{j_R} = 0$, since the geodesic connecting points $T = (t_T, \mathbf{r}_T)$ and $R = (t, \mathbf{r})$ is null.

A world line is an invariant geometric quantity [23]. For the limiting case of a satellite in Minkowski space-time, the time component t_s of the satellite world line in Eq. (35) defines a scalar field, $t_s = t_s(t, \mathbf{r})$, that is given by the implicit equation

$$t_{s}(t,\mathbf{r}) = t - \frac{1}{c} \left| \mathbf{r} - \mathbf{r}_{s}(t_{s}(t,\mathbf{r})) \right|$$
(37)

where $\mathbf{r}_s(t_s)$ is the ephemeris of the satellite. It is easy to check that this particular form for $t_s(t, \mathbf{r})$ in the phase function $\Phi_s(t, \mathbf{r})$ in Eq. (35) satisfies the eikonal Eq. (36).

For the actual case where we take into account the Earth's gravitational field, the light cone equation is given by $\Omega(T,R)=0$, where Ω is the world function given in Eq. (28). Writing the world function in Eq. (28) in the form

$$\Omega = -\frac{1}{2}(1+\alpha)(x^0 - x_s^0)^2 + \frac{1}{2}(1+\beta)(\mathbf{r} - \mathbf{r}_s)^2 \quad (38)$$

where α and β are small quantities, the scalar phase field $\Phi_s(\mathbf{r}, t)$ in Eq. (35) can be written in terms of the function $x_s^0(x^0, \mathbf{r})$ that is implicitly given by

$$x_{s}^{0}(x^{0},\mathbf{r}) = x^{0} - \left(1 + \frac{1}{2}(\beta - \alpha)\right) |\mathbf{r} - \mathbf{r}_{s}(x_{s}^{0}(x^{0},\mathbf{r}))| \quad (39)$$

where we have kept only linear terms in α and β . Note that α and β are two-point functions that depend on space-time points $T = (x_s^0, \mathbf{r}_s)$ and $R = (x^0, \mathbf{r})$.

An event in this space-time can be uniquely labeled by four real hardware times, $(\tau_1^*, \tau_2^*, \tau_3^*, \tau_4^*)$, or alternatively by dimensionless real numbers, $(\eta_1, \eta_2, \eta_3, \eta_4)$, where each η_s gives the continuous code parameter at emission time for satellite *s*. This system of coordinates has been studied by Synge who called them optical coordinates [23]. More recently, these same coordinates have been called GPS coordinates and their theoretical properties of have been explored in some detail [13,14].

At any time, a GPS user has more than four satellites in view, so in the real implementation of the GPS system of coordinates, they are multiple valued. Since the satellite signals are line of sight, a GPS user sees satellites rise and set on the horizon, and a different set of four satellites defines the 3D hypersurfaces. There are currently approximately 24 satellites in the GPS constellation, and 37 code sequence possibilities [21], so at most, the numbers $s = 1,2,3, \ldots, 37$.

In our discussion, we have neglected atmospheric effects. In practice, for users of GPS in the Earth's atmosphere, there exist significant propagation delays as well as frequency dispersive effects. The Earth's troposphere (atmosphere from the ground to approximately 10 km) causes the same time delay for both L_1 and L_2 frequencies. On the other hand, the Earth's ionosphere (60 km to 700 km altitude) is dispersive at these frequencies, due to the presence of free electrons,



FIG. 4. A mechanical analogue of a GPS receiver tracks the code broadcast from a GPS satellite. The receiver has two code wheels. The phase angle of code wheel 1 represents the current time in the receiver. The phase angle of code wheel 2 is kept aligned with the incoming code sequence from the satellite.

and hence causes two different delays for the L_1 and L_2 frequencies. A first principles treatment is possible, following Synge [23]. However, in practice, simpler methods are used for correcting for both of these effects in GPS receivers [21]. Atmospheric effects will not be considered further here. However, we note that signal propagation through the (frequency dispersive) atmosphere leads to a removal of the discontinuity in the electromagnetic tensor F_{ij} . Hence, the dispersive effects of the atmosphere limit the accuracy of GPS for a ground user because the code edge is not sharp after signal propagation. The magnitude of this effect has not been investigated. Also, the finite size of the transmitting antenna also leads to a code edge that is not sharp [15].

VI. PSEUDORANGE MEASUREMENT

A GPS receiver makes a special type of measurement, called a pseudorange measurement, in which the Doppler and gravitational frequency shifts are essentially removed (up to an additive constant) from the measurement. A GPS receiver internally replicates (generates) the satellite P-code s(t) for each satellite. The receiver determines which satellite signal it is receiving by matching the internally replicated P-code with the code received from the satellite. In what follows, we describe the pseudorange measurement made by a GPS receiver by using a mechanical analogue. Our mechanical analogue of a GPS receiver has two code wheels, numbered 1 and 2, see Fig. 4. During operation of the receiver, both code wheels advance at the same angular velocity, Ω_R . The receiver will align the code incoming from the satellite with the code on wheel 2. The code arriving is like a

bicycle chain that is to be placed onto the sprocket (code wheel 2). In order to align the codes, the receiver must match the rate (frequency) at which bits are arriving as well as match the duration of each bit. In general, the receiver has a non-zero velocity in the ECI frame and it is at a different gravitational potential than the transmitting satellite. At any instant in the receiver's comoving frame, the frequency of code bits received from the satellite is $f_R = \omega_R/2\pi$, where f_R is the number of received bits per second of the code $s(\tau^*)$. To align the codes, this frequency ω_R must be made equal to ω_o , which is the frequency seen in the comoving frame of the receiver, given in Eq. (20).

The temporal duration of the bits on code wheel 2 and the temporal duration of incoming bits from the satellite are made equal by adjusting the radius r_w of code wheel 2. Code wheel 1 is adjusted to the same radius as code wheel 2. If the code wheel has a circumference equal to $2\pi r_w$, the length per bit along the the circumference is $2\pi r_w/N$, where there are N bits in the code. In order to match incoming bits, the code wheel must have a linear velocity $2\pi r_w f_R/N = r_w \Omega_R$. The angular velocity of the code wheel is then $\Omega_R = \omega_R/N$. This angular velocity of the code wheel, and consequently the frequency at which the replicated code is generated, ω_R , will be adjusted continually and depends on the motion of the receiver and satellite, so that the replicated code is constantly aligned with the code arriving from the satellite:

$$\omega_R = \omega_o \,. \tag{40}$$

Therefore, a GPS receiver searches in time (aligns code phase on wheel 2 with incoming code) and in frequency (adjustment of r_w to make temporal length of bits on wheel 2 equal to satellite bits). This alignment of the code is commonly referred to as "code tracking" or "code lock."

At any instant in time, code wheel 1 has a current value of phase, Φ_R , which is indicated by the pointer, see Fig. 4. The value of the replicated phase Φ_R on code wheel 1 is a representation of the receiver's current hardware time, τ_R^* . The phase Φ_R and receiver clock hardware time τ_R^* are related by

$$\Phi_R = \frac{\omega_s}{N} \tau_R^* \tag{41}$$

where ω_s is a system constant given by Eq. (18). The system constant ω_s is a conversion factor from hardware time, τ_R^* , to phase Φ_R . A key point is that during "code tracking" the phase Φ_R on code wheel 2 is adjusted to keep pace with the incoming code from the satellite, but this is done by advancing the time τ_R^* , and not by changing the frequency ω_s , which is a fixed constant in the system.

The replicated phase in the receiver can be related to the receiver's world line by writing the hardware time τ_R^* in terms of the coordinate time along the receiver's world line in the ECI frame, x_R^0 ,

$$x_R^0 = c \left(\tau_R^* + \Delta \tau_R^* \right) \tag{42}$$

where $\Delta \tau_R^*$ is the time correction to the receiver's hardware time, commonly called a clock bias. The receiver time, $\tau_R^* = \eta T/N$, is periodic, $\tau_R^* + T = \tau_R^*$, where T = 1 week [the period of the P-code $s(\tau_R^*)$], and η is a real number $0 \le \eta \le N$, which corresponds to the interpolated value of the bit number *n* in the P-code $s(\tau_R^*)$, see Fig. 4. This periodicity in τ_R^* does not introduce any ambiguity because it is assumed that a GPS user is located within one light week distance from the Earth. Using Eq. (42) the receiver phase, Φ_R , can be related to the geometry by

$$\Phi_R = \frac{\omega_s}{cN} (x_R^0 - c\Delta \tau_R^*).$$
(43)

On code wheel 2, the receiver has the replicated code (for a given satellite). This replicated code is rotated back and forth on code wheel 2 until it is aligned with the code phase $\Phi_s(t_R, \mathbf{r}_R)$ arriving from the satellite, where \mathbf{r}_R is the position of the receiver. As described above, at all times the phase on code wheel 2 is aligned with the incoming code phase, in both time (angle of the wheel) and frequency or bit duration (radius r_w is adjusted).

The pseudorange measurement made by a GPS receiver at space-time point (t_R, \mathbf{r}_R) is the difference of the phases on code wheel 1 and 2 [30]

$$\Delta \Phi(t_R, \mathbf{r}_R) \equiv \Phi_R - \Phi_s(t_R, \mathbf{r}_R) \tag{44}$$

where Φ_R is the curent value of the replicated code phase in the receiver (on code wheel 1) and $\Phi_s(t_R, \mathbf{r}_R)$ is the value of the broadcast phase $\Phi(t, \mathbf{r})$ for satellite *s* [given in Eq. (35)] evaluated at the receiver at space-time measurement point (t_R, \mathbf{r}_R) . The phase difference, $\Delta \Phi(t_R, \mathbf{r}_R)$, is a bonafide measurement [29] because it is a comparison (a difference) between the value of the scalar field at the receiver, $\Phi_s(t_R, \mathbf{r}_R)$, and the replicated code phase, Φ_R . The measured difference, $\Delta \Phi(t_R, \mathbf{r}_R)$, is commonly expressed in units of length and is called a pseudorange measurement from receiver to satellite *s*:

$$\rho_s = c \frac{N}{\omega_s} [\Phi_R - \Phi_s(t_R, \mathbf{r}_R)]$$
(45)

where ρ_s is the measured pseudorange in units of length.

The measured pseudorange, ρ_s , is a scalar under generalized coordinate transformations (e.g., under Lorentz transformations) because the phase function $\Phi_s(t, \mathbf{r})$ is a scalar field and Φ_R is an invariant that depends on the world line of the receiver. The pseudorange depends on both the world line of the satellite and the world line of the receiver, which are invariant quantities. The world line of satellite $s, x_s^i(\tau^*)$, enters into the definition of the scalar field $\Phi_s(t,\mathbf{r})$, see Eq. (35). The world line of the receiver enters into the value of the phase of the replicated code, Φ_R , see Eq. (43). The world line of the receiver also enters into the evaluation of the scalar field $\Phi_s(t,\mathbf{r})$ at the position of the receiver at event $R = (t_R, \mathbf{r}_R)$. Therefore, the pseudorange measurement, ρ_s , is a two-point scalar that depends on the emission event T and the reception event R. The pseudorange is a particular case of quantities known as two-point tensors, which depend on two space-time points, and have tensorial transformation properties with respect to generalized coordinate transformations (that can differ) at each point [23]. Therefore, the pseudorange ρ_s is a two-point scalar under separate generalized coordinate transformations at $T = (t_T, \mathbf{r}_T)$ and at $R = (t_R, \mathbf{r}_R)$ and can be labeled as $\rho(t_T, \mathbf{r}_T, t_R, \mathbf{r}_R)$ [23]. The world function of space-time, $\Omega(x_1^i, x_2^i)$ given in Eq. (28), is a wellknown two-point scalar, and not surprisingly, the pseudorange and the world function have the same transformation properties.

Two-point scalar fields enter into measurement situations whenever there is a field generated at space-time point T, such as an electromagnetic field, and a measurement or *projection* is done at the measurement event at space-time point R. In this sense, there is some similarity between a relativistic treatment of measurement processes and a quantum mechanical treatment, which is mentioned in the Introduction. In the case of relativity, a measurement of a field is a projection of the field onto an observer's tetrad [23,31]. In quantum mechanics, traditionally projection operators are invoked in the definition of measurement [32–34].

VII. TIME TRANSFER AND NAVIGATION IN SPACE-TIME

The measured pseudorange, ρ_s in Eq. (45), can be related to the geometric range between the emission event $T = (t_T, \mathbf{r}_T)$ and the reception event $R = (t_R, \mathbf{r}_R)$ by using Eq. (43) for Φ_R and evaluating the broadcast phase $\Phi(t, \mathbf{r})$ for satellite *s* [given in Eq. (35)] at the reception event *R*, leading to

$$\rho_s = |\mathbf{r}_R - \mathbf{r}_T| + c \Delta \tau_T^* - c \Delta \tau_R^* + \Delta(\mathbf{r}_T, \mathbf{r}_R)$$
(46)

where we have set $\Delta \tau_T^* = \Delta \tau_\eta^*$. In Eq. (46), $|\mathbf{r}_R - \mathbf{r}_T|$ is the geometric range between events *T* and *R*, $\Delta \tau_R^*$ is the clock correction to the receiver clock at event *R*, and $c\Delta \tau_T^* = x_s^0(T) - c\tau_s^*(T)$ is the satellite clock correction at event *T*. The satellite clock correction, $\Delta \tau_T^*$, is broadcast by the satellite in the navigation message contained in the function $D(\tau^*)$, see Eq. (23). Conventionally, the pseudorange is modelled [9–11] in flat space-time as $\rho_s = |\mathbf{r}_R - \mathbf{r}_T| + c\Delta \tau_T^* - c\Delta \tau_R^*$. The term $\Delta(\mathbf{r}_T, \mathbf{r}_R)$ is a small correction due to the presence of the Earth's gravitational field that modifies space-time geometry near Earth. This correction depends on the mass of the Earth, *M*, the angular velocity of Earth rotation, Ω , and the Earth's equatorial radius, *R*, and is given by

$$\Delta(\mathbf{r}_T, \mathbf{r}_R) = \frac{2GM}{c^2} \left(\Lambda(\mathbf{r}_T, \mathbf{r}_R) - \frac{|\mathbf{r}_R - \mathbf{r}_T|}{R} \right) - \frac{\Omega^2 R^2}{c^2} |\mathbf{r}_R - \mathbf{r}_T|$$
(47)

where the purely geometric function $\Lambda(\mathbf{r}_T, \mathbf{r}_R)$ is given in Eq. (29). The quantity $2GM/c^2 \approx 0.89$ cm is the gravitational radius of the Earth, and $\Omega^2 R^2 |\mathbf{r}_R - \mathbf{r}_T|/c^2 \approx 4.8 \times 10^{-5}$ m, using $|\mathbf{r}_R - \mathbf{r}_T| \approx a - R$, where *a* is the semimajor axis of the

GPS satellites. The pseudorange in Eq. (46) is based on the metric in the ECI frame of reference, which is given in Eq. (10).

If a user of the GPS knows their spatial coordinates in the ECI frame, \mathbf{r}_R , then a single pseudorange measurement ρ_s to one satellite is sufficient to determine the users time t_R from Eq. (46). In Eq. (46) $\mathbf{r}_T = \mathbf{r}_s(t_T)$, where $\mathbf{r}_s(t_T)$ is the broadcast satellite ephemeris evaluated at transmit time t_T that satisfies

$$\Omega(t_T, \mathbf{r}_s(t_T), t_R, \mathbf{r}_R) = 0 \tag{48}$$

where $\Omega(t_T, \mathbf{r}_T, t_R, \mathbf{r}_R)$ is the world function given in Eq. (28). Therefore, Eqs. (46) and (48) can be solved numerically for the user's reception event time t_R , which is a coordinate time in the ECI metric in Eq. (10).

The more common case is that a user of the GPS wants to obtain time but he knows his spatial coordinates in the ECEF frame of reference, $\mathbf{y} = (y^1, y^2, y^3)$. The user's ECI frame coordinates, $\mathbf{r}_R = (x^1, x^2, x^3)$, are given by

$$\mathbf{r}_R = D(t_R - t_0) \cdot \mathbf{y} \tag{49}$$

where $D(t-t_0)$ is a time-dependent rotation matrix, which is equal to the unit matrix at epoch time $t=t_0$, see Eq. (3). In this case, to obtain the user's time t_R at the reception event, \mathbf{r}_R in Eqs. (46) and (48) must be eliminated by use of Eq. (49). Equations (46) and (48) can be solved for t_R .

Navigation in space-time means the simultaneous determination of the user position and time, i.e., determination of the space-time coordinates of the reception event $R = (t_R, \mathbf{r}_R)$. Navigation can be carried out by simultaneously measuring four pseudoranges, ρ_s , to four different satellites, s=1,2,3,4. It is clear that four simultaneous equations of the form in Eq. (46) can be solved for the four coordinates (t_R, \mathbf{r}_R) that specify the user reception event in the ECI frame. Typically, a GPS user on Earth wants to know their coordinates in the ECEF frame of reference. The user must determine the coordinate time of emission for one satellite from Eq. (48) and use it in the transformation Eq. (49) to determine user ECEF coordinates $\mathbf{y}=(y^1, y^2, y^3)$ [35].

VIII. RESOLUTION OF THE TWO RECEIVER EXPERIMENT

If two receivers are at the same event M in space-time, and they are tracking the same four satellites, do the receiver's measure the same pseudorange? Also, do they obtain the correct space-time coordinates for the event M? The resolution of the two receiver thought experiment, which was described in Sec. II, is now clear. The value of the phase received from each satellite, $\Phi_s(t_R, \mathbf{r}_R)$, is the same for each receiver, independent of their velocities, because $\Phi_s(t, \mathbf{r})$ is a scalar field. However, the value of the hardware time for each receiver, and therefore the current replicated code phase, Φ_R , is different for each receiver. This difference is due to the different clock bias $\Delta \tau_R^*$ of the two receivers. We can then say that each measured pseudorange is an invariant under coordinate transformations; however, the pseudoranges (to the same satellite) are different for each receiver because they depend on receiver world lines. For navigation purposes, even though the set of pseudoranges are different for each of the two receivers, the clock bias for each receiver is determined in the course of solving the four equations in Eqs. (46) for s=1,2,3,4. Therefore, while each receiver obtains a *different* set of four invariant pseudoranges to the same four satellites, each receiver obtains the correct spacetime coordinates of the reception event $M = (t_R, \mathbf{r}_R)$.

IX. SUMMARY

We used a two-receiver thought experiment in Sec. II to motivate the need to understand the transformation properties of the measured pseudorange; the quantity that is measured in the GPS.

Starting from the weak field approximation for the metric of space-time in the vicinity of the Earth, we used the standard transformation to the rotating frame, and defined a new coordinate time so that coordinate time on the geoid surface is equal to proper time. Then we transform from the ECEF frame to the ECI frame, to obtain the metric that describes the ECI frame of reference near the Earth, given in Eq. (10). In this metric, due to a combination of gravitational and time dilation effects, the GPS satellite clocks appear to run fast with respect to coordinate time. We described the "factory offset" that is routinely applied to slow down the satellite clocks to (approximately) keep coordinate time in the ECI frame. In Sec. IV B, for the metric in the ECI frame given in Eq. (10), we derived the apparent frequency of the satellite signal seen by an observer at an arbitrary position moving with an arbitrary velocity, due to Doppler and gravitational potential differences, see Eq. (20).

We described the nature of the digital signal (P-code) that is broadcast by GPS satellites, and the space-time grid created by GPS satellites due to the discontinuities in the broadcast electromagnetic field. We computed the world function of space-time, given in Eq. (28), for the ECI metric in Eq. (10). Using this world function, we defined a scalar phase field $\Phi_s(t,\mathbf{r})$, given in Eq. (35), that satisfies the eikonal equation. This phase field is seen in the same way by all observers, independent of their state of motion. Using this phase field, and a mechanical analogue for a GPS receiver, we defined the measured pseudorange in Eq. (45), and found that under generalized coordinate transformations it transforms as a two-point scalar, just like the world function. Within the geometrical optics approximation, there are no velocity effects on pseudorange measurements. Finally, we related the measured pseudorange to the geometry of spacetime in Eq. (46). We obtained a small correction, given by $\Delta(\mathbf{r}_T, \mathbf{r}_R)$ in Eq. (47), to the conventional model of the pseudorange. This correction is due to the curvature of space-time in the vicinity of the Earth.

ACKNOWLEDGMENTS

This work was supported by the Advanced Research and Development Activity (ARDA).

APPENDIX A: CONVENTIONS AND NOTATION

We use the convention that roman indices, such as found on space-time coordinates x^i , take the values i=0,1,2,3, and greek indices take values $\alpha = 1,2,3$. Summation is implied over the range of an index when the same index appears in a lower and upper position. If x^i and $x^i + dx^i$ are two events along the world line of an ideal clock, then the proper time interval between these events is $d\tau = ds/c$, where ds is given in terms of the space-time metric as $ds^2 = -g_{ij} dx^i dx^j$. We choose g_{ij} to have the signature +2, so when g_{ij} is diagonalized at any given space-time point, the elements can take the form of the Minkowski metric given by $\eta_{00} = -1$, $\eta_{\alpha\beta} = \delta_{\alpha\beta}$.

APPENDIX B: DOPPLER EFFECT IN EARTH'S GRAVITATIONAL FIELD

A satellite at spatial position \mathbf{r}_s and travelling at velocity \mathbf{v}_s emits a signal in its comoving frame with frequency $\boldsymbol{\omega}_s$. An observer at spatial position \mathbf{r}_o and travelling at velocity \mathbf{v}_o measures the frequency of this signal to be $\boldsymbol{\omega}_o$ in his comoving frame. These frequencies are related by Eq. (20). This relation can be derived by evaluating [23]

$$\frac{\omega_o}{\omega_s} = 1 - \frac{\Omega_{i_s} V^{i_s} + \Omega_{i_o} V^{i_o}}{\Omega_i V^{i_s}} \tag{B1}$$

where $\Omega_{i_{a}}$ and $\Omega_{i_{a}}$ are the covariant derivatives of the world function with respect to satellite and receiver coordinates, x_s^i and x_o^j , respectively, and V^{i_s} and V^{i_o} are the components of the 4-velocities of the satellite and the receiver at space-time points $S = x_s^j = (t_T, \mathbf{r}_T)$ and $O = x_o^i = (t_R, \mathbf{r}_R)$, respectively. The form of Eq. (B1), together with the transformation properties of the world function, shows that the quantity $\omega_{\alpha}/\omega_{s}$ is a two-point scalar under generalized coordinate transformations. We have computed the world function in Eq. (B1), $\Omega = \Omega(x_s^i, x_o^j)$, and it is given in Eq. (28) for the metric given in Eq. (10). To obtain an explicit expression for ω_o/ω_s we use an alternative derivation [23]. Consider the satellite and observer world lines, $x_s^{l}(s_s)$ and $x_a^{l}(s_a)$, which are parametrized by proper times s_s/c and s_o/c , respectively, see Fig. 5. The lines AB and CD are null geodesics that connect satellite and receiver world lines, at successive cycles of the emitted signal.



FIG. 5. The world line of satellite and receiver are shown together with the 3D hypersurfaces of constant phase between points AB and CD.

The frequency ratio ω_o/ω_s is related to the proper times

$$\frac{\omega_o}{\omega_s} = \frac{ds_s}{ds_o} \tag{B2}$$

where ds_s is the proper time between AC and ds_o is the proper time between BD. Using the metric in Eq. (10), the ratio of frequencies can be written as

$$\frac{\omega_o}{\omega_s} = \frac{ds_s}{ds_o} = \left[\frac{g_{00}(S) + g_{\alpha\beta}(S)v_s^{\alpha}v_s^{\beta}}{g_{00}(O) + g_{\alpha\beta}(O)v_o^{\alpha}v_o^{\beta}}\right]^{1/2} \frac{\Delta x_s^0}{\Delta x_o^0} \quad (B3)$$

where the emission and reception events are connected by a null geodesic, so that $\Omega(S,O) = 0$. In Eq. (B3), $g_{\alpha\beta}(S)$ is the metric on the satellite world line at event *S*, $g_{\alpha\beta}(O)$ is the metric on the observer world line at event *O*, and v_s^{α} and v_o^{α} are the velocity components of satellite and receiver, for α =1,2,3. For example, the satellite velocity is $v_s^{\alpha} = \Delta x_s^{\alpha}/\Delta x_s^0$, where Δx_s^0 is the coordinate time between events *AC* on the satellite world line. In the limit $\Delta x_s^0 \rightarrow 0$ (and therefore $\Delta x_o^0 \rightarrow 0$), using the world function in Eq. (28) to connect points *AC* and *BD*, a lengthy calculation of Eq. (B3) using the metric in Eq. (10) leads to Eq. (20). The term $\Delta x_s^0/\Delta x_o^0$ in Eq. (B3) leads to linear velocity terms in Eq. (20).

- [1] I. L. Chuang, Phys. Rev. Lett. **85**, 2006 (2000); quant-ph/0004105.
- [2] R. Jozsa, D. S. Abrams, J. P. Dowling, and C. P. Williams, Phys. Rev. Lett. 85, 2006 (2000).
- [3] J. Preskill, quant-ph/0010098.
- [4] V. Giovannetti, S. Lloyd, and L. Maccone, Nature (London) 412, 417 (2001).
- [5] Y. Shih (unpublished).
- [6] E. A. Burt, C. R. Ekstrom, and T. B. Swanson, Phys. Rev. Lett.

87, 129801 (2001); "A Reply to "Quantum Clock Synchronization," quant-ph/0007030.

- [7] R. Jozsa, D. S. Abrams, J. P. Dowling, and C. P. Williams, Phys. Rev. Lett. 87, 129802 (2001).
- [8] U. Yurtsever and J. P. Dowling, Phys. Rev. A 65, 052317 (2002).
- [9] Global Positioning System: Theory and Applications, Vols. I and II, edited by B. W. Parkinson and J. J. Spilker, Progress in Astronautics and Aeronautics, Vols. 163 and 164 (Amer. Inst.

Aero. Astro., Washington, D.C., 1996).

- [10] E. D. Kaplan, Understanding GPS: Principles and Applications, Mobile Communications Series (Artech House, Boston, 1996).
- [11] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, *Global Positioning System Theory and Practice* (Springer-Verlag, New York, 1993).
- [12] The Russian satellite system known as GLONASS is similar to GPS.
- [13] C. Rovelli, Phys. Rev. D 65, 044017 (2002).
- [14] M. Blagojevi, J. Garecki, F. W. Hehl, and Yu. N. Obukhov, Phys. Rev. D 65, 044018 (2002).
- [15] A. H. Krall and T. B. Bahder, J. Appl. Phys. 90, 6513 (2001).
- [16] Actually there are three main relativistic effects that contribute significantly, see Sec. II.
- [17] L. D. Landau and E. M. Lifshitz, *Classical Theory of Fields* (Pergamon, New York, Fourth Revised English edition, 1975).
- [18] M. Caputo, *The Gravity Field of the Earth* (Academic, New York, 1967).
- [19] N. Ashby and J. J. Spilker, in *Global Positioning System: Theory and Applications*, Vols. I and II, edited by B. W. Parkinson and J. J. Spilker, Progress in Astronautics and Aeronautics, Vols. 163 and 164 (Amer. Inst. Aero. Astro., Washington, D.C., 1996).
- [20] R. S. Nerem, F. J. Lerch, J. A. Marshall, E. C. Pavlis, B. H. Putney, B. D. Tapley, R. J. Eanes, J. C. Ries, B. E. Schutz, C. K. Shum, M. M. Watkins, S. M. Klosko, J. C. Chan, S. B. Luthcke, G. B. Patel, N. K. Pavlis, R. G. Williamson, R. H. Rapp, R. Biancale, and F. Nouel, J. Geophys. Res. 99, 24421 (1994b).
- [21] See J. J. Spilker, in *Global Positioning System: Theory and Applications*, Vols. I and II [9], Chap. 2.
- [22] Anonymous, "Navstar GPS Space Segment/Navigation User Interfaces," Revision C, Initial Release, ARINC Research Corporation, 10 October 1993, ICD-GPS-200.

- [23] J. L. Synge, *Relativity: The General Theory* (North-Holland, New York, 1960).
- [24] A. Kheyfets, "Spacetime geodesy," Weapons Laboratory, Air Force Systems Command, Kirtland Air Force Base, New Mexico, Technical Report No. WL-TN-90-13, 1991.
- [25] This statement neglects the correction needed due to the eccentricity of the satellite orbit, the so-called "*e* sin *E* effect."
- [26] V. Fock, *The Theory of Space, Time and Gravitation* (Pergamon, Oxford, England, 1964).
- [27] F. De Felice and C. J. S. Clarke, *Relativity on Curved Manifolds*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, New York, 1990).
- [28] T. B. Bahder, Am. J. Phys. 69, 315 (2001).
- [29] The evaluation of a scalar field $\Phi_s(t, \mathbf{r})$ at a point of observation (at the receiver), (t_R, \mathbf{r}_R) , is the same measurement operation as the projection of a tensor field F_{ij} onto an observer tetrad $\lambda_{(a)}^i$, where *i* labels the coordinate vector components and a = 1, 2, 3, 4 labels the tetrad basis vectors. For the tensor field, the projection is done by the formula $F_{(ab)} = F_{ij} \lambda_{(a)}^i \lambda_{(a)}^j$, where $F_{(ab)}$ are invariants under coordinate transformations, but they depend on the observer world line through the definition of the tetrad, see Ref. [23].
- [30] P. Ward, in E. D. Kaplan, Understanding GPS: Principles and Applications, Mobile Communications Series (Artech House, Boston, 1996), Chap. 5.
- [31] F. A. E. Pirani, Bull. Acad. Pol. D. Sci. Cl. III V, 143 (1957).
- [32] J. von Neumann, Mathematical Foundations of Quantum Mechanics (Princeton University Press, Princeton, 1955).
- [33] R. Omnes, *The Interpretation of Quantum Mechanics* (Princeton University Press, Princeton, 1994).
- [34] See also the discussion in J. Preskill, Lecture Notes for Physics 229: Quantum Information and Computation (California Institute of Technology, Pasadena, CA, 1998), http:// www.theory.caltech.edu/people/preskill/ph229/
- [35] N. Ashby and M. Weiss, *Global Positioning System Receivers* and *Relativity*, NIST Technical Note 1385 (U.S. GPO Washington, D.C. 1999).