Feature Article

Relativity and the Global Positioning System

We need general relativity to understand extreme astrophysical realms. But the theory also turns out to be essential for the many mundane activities that nowadays rely on the precision of the GPS.

Neil Ashby

It's been almost a century since Einstein introduced the special theory of relativity. All observational tests to date confirm both the special and the general theory. These tests have ranged from sensitive laboratory experiments involving optics, atoms, nuclei, and subnuclear particles to the observation of orbiting clocks, planets, and objects far beyond the Solar System.

The general theory of relativity will soon be tested with high precision by Stanford University's Satellite Test of the Equivalence Principle (STEP),[1] and observations by the worldwide array of gravitational-wave detectors presently under construction are expected to test the theory in the extreme realm of strong gravitational fields and high velocities (see the articles by Clifford Will and by Barry Barish and Rainer Weiss in Physics Today, October 1999, pages 38* and 44*, respectively).

Numerous relativistic issues and effects play roles in the global positioning system, on which millions of drivers, hikers, sailors, and pilots depend to find out where they are. The GPS system is, in effect, a realization of Einstein's view of space and time. Indeed, the system cannot function properly without taking account of fundamental relativistic principles. That is the subject of this article.

The global positioning system

The orbiting component of the GPS consists of 24 satellites (plus spares): four satellites in each of six different planes inclined 55° from Earth's equatorial plane. The satellites are positioned within their planes so that, from almost any place on Earth, at least four are above the horizon at any time. Orbiting about 20 000 km above Earth's surface, all the satellites have periods of 11 hours and 58 minutes. Because that's half a sidereal day, a fixed observer on the ground will see a given satellite at almost exactly the same place on the celestial sphere twice each day. Each satellite carries one or more very stable atomic clocks, so that the satellites can transmit synchronous timing signals. The signals carry coded information about the transmission time and position of the satellite.

Figure 1 shows one of the new generation of GPS orbiters (called Block IIR satellites) that have recently begun replacing the older generation. Its antenna array efficiently beams right-circularly polarized radio signals toward Earth's surface. By the time the spreading radio signal reaches the ground, its intensity is only about 3 X 10^{-14} W/m^2. To process such faint signals, GPS receivers must implement very special techniques. Figure 2 shows the new Block IIR satellites on an assembly line.

Figure 1

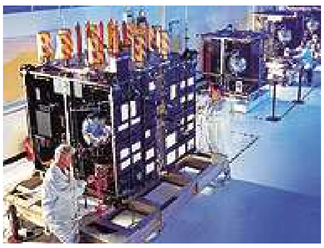

Figure 2

Data transmitted by the satellites are continuously monitored by receiving stations around the globe and forwarded to a master control station, where satellite orbits and clock performance are computed. The resulting orbital and clock data are then uploaded to the satellites for retransmission to users.

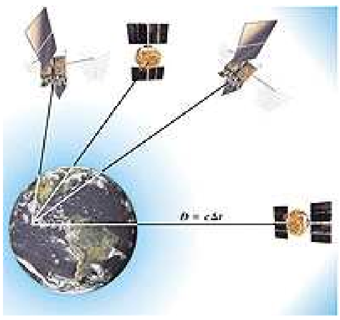

Figure 3

The fundamental principle on which GPS navigation works is an apparently simple application of the second postulate of special relativity--namely, the constancy of c, the speed of light. Referring to figure 3, suppose that a receiver, on or near Earth's surface, simultaneously receives signal pulses from four satellites, transmitted at times t_i from satellites at positions $\mathbf{r}_i$. Then the position $\mathbf{r}$ of the receiver and the time t on its clock when the four signals arrive can be determined by solving four simultaneous equations

$$|\mathbf{r} - \mathbf{r}_i| = c(t - t_i);\ i = 1, 2, 3, 4. \quad (1)$$

These propagation-delay equations, strictly valid in an inertial frame, are the basis for position and time determination by the GPS receivers.

Also This Month
Quantum Cascade Lasers
Relativity and the Global Positioning System
Advanced Physics in the High Schools

Related readings from *Physics Today*
October 1981, page 20
December 1993, page 17
October 1999, pages 38
October 1999, page 44
January 1994, page 9
April 1993, page 9
March 2001, page 37

Accurate navigation with the GPS is made possible by the phenomenal performance of modern atomic clocks.[2] If navigation errors of more than a meter are to be avoided, an atomic clock must deviate by less than about 4 nanoseconds from perfect synchronization with the other satellite clocks. That amounts to a fractional time stability of better than a part in 10^{13}. Only atomic clocks can do that. Even so, the system requires frequent uploads of clock corrections to the satellites.

The reference for GPS time is a composite clock based on the US Naval Observatory's ensemble of about 50 cesium-beam frequency standards and a dozen hydrogen masers. Clock times on GPS satellites usually agree with the observatory's ensemble to within about 20 ns.

Relativistic effects are much larger than a part in 10^{13}. For example, satellite speeds v are about 4 km/s. Time dilation then causes the moving clocks' frequencies to be slow by Df/f = $v^2/2c^2$» 10^{-10}. Gravitational effects are even larger. In fact, relativistic effects are about 10 000 times too large to ignore.

Suppose one wanted to improve GPS spatial precision so that receiver positions could be determined with an uncertainty of only a centimeter. A radio wave travels 1 cm in 0.03 ns. So one would have to account for all temporal relativistic effects down to a few hundredths of a nanosecond. But the second-order Doppler shift of an orbiting atomic clock, if it were not taken into account, would cause an error this large to build up in less than half a second. An effect of comparable size is contributed by the gravitational blueshift, which results when a photon--or a clock--moves to lower altitude. If these relativistic effects were not corrected for, satellite clock errors building up in just one day would cause navigational errors of more than 11 km, quickly rendering the system useless.

Self-consistent synchronization

Clocks moving along different trajectories in space and on Earth undergo different gravitational and motional frequency shifts. The "proper times" recorded by all these clocks in their own rest frames quickly diverge. Therefore one needs some reasonable means of synchronization, in order that equations 1 have their intended meaning--expressing signal propagation at speed c in straight lines in an inertial frame. The times t_i at which the transmissions originate must be established by a self-consistent synchronization scheme.

In Earth's neighborhood, the field equations of general relativity involve only a single overall time variable. While there is freedom in the theory to make arbitrary coordinate transformations, the simplest approach is to use an approximate solution of the field equations in which Earth's mass gives rise to small corrections to the simple Minkowski metric of special relativity, and to choose coordinate axes originating at the planet's center of mass and pointing toward fixed stars. In this Earth-centered inertial (ECI) reference frame, one can safely ignore relativistic effects due to Thomas precession or Lense-Thirring drag. The gravitational effects on clock frequency, in this frame, are due to Earth's mass and its multipole moments.

In the ECI frame, the fundamental invariant spacetime interval ds^2 of general relativity can be written in the approximate form

$$\mathrm{d}s^2 = -\left(1+\frac{2\Phi}{c^2}\right)(c\ \mathrm{d}t')^2 + \left(1-\frac{2\Phi}{c^2}\right)(\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2)$$

, (2)

where F < 0 is the Newtonian gravitational potential. For the GPS, we can ignore terms of order smaller than c^{-2}. The variable t' in the equation is called the coordinate time. In general relativity, one can construct a consistent spacetime coordinate system for a "patch" that encompasses Earth and its GPS satellites without having to resort to more than the one such time variable. One can think of this coordinate time as the proper time on an atomic clock at rest far away from Earth's gravity.

However, the rate of International Atomic Time (TAI) is based on atomic clocks resting essentially at sea level, where they are subject to second-order Doppler shifts due to Earth's rotation and gravitational redshifts relative to clocks 20 000 km higher up. The two different time variables can be reconciled by scaling the rate of coordinate time so that it matches the rate of TAI. The time variable t actually used in the GPS is related to the coordinate time t' of equation 2 by $t' = t(1 - U/c^2)$, where the constant parameter U includes motional effects due to Earth's rotation and gravitational effects from its mass distribution.

It is very useful that Earth's geoid--the planet's idealized sea-level surface--is a surface of constant effective gravitational potential U in an Earth-fixed rotating reference frame, so that all atomic clocks at rest on the geoid tick at the same rate. That's a nontrivial consequence of a combination of effects arising from time dilation and the multipole expansion of Earth's nonspherical mass distribution.[3,4] To an approximation good enough for the GPS, the constant U can be calculated in terms of Earth's mass, its quadrupole moment, and its rotational angular velocity W_E. Then the metric can be written as

$$\mathrm{d}s^2 = -\left(1+\frac{2(\Phi-U)}{c^2}\right)(c\ \mathrm{d}t')^2 + \left(1-\frac{2\Phi}{c^2}\right)(\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2)$$

, (3)

and the proper rate of all atomic clocks at rest on the geoid will be given by dt = ds/c.

For an atomic clock moving along some arbitrary path, one can envision measuring the clock's proper time increment ds/c, solving equation 3 for dt, and then integrating dt along the path to get the elapsed coordinate time *t*. Thus, for each atomic clock, the GPS generates a "paper clock" that reads *t*. All coordinate clocks generated in this way would be self-consistently synchronized if one brought them together--assuming that general relativity is correct. That, in essence, is the procedure used in the GPS.[3,4]

In equation 3, the leading contribution to the gravitational potential F is the simple Newtonian term $-GM_E/r$. The picture is Earth-centered, and it neglects the presence of other Solar-system bodies such as the Moon and Sun. That they can be neglected by an observer sufficiently close to Earth is a manifestation of general relativity's equivalence principle.[5]

In the ECI frame, the only detectable effects of distant masses are their residual tidal potentials. Tidal effects on orbiting GPS clocks due to the Moon and Sun amount to less than a part in 10^{15}. Currently they are ignored. But tidal potentials do have a significant effect on satellite orbits.

GPS receivers

The GPS system transfers transmission coordinate times t_i to a receiver in a very sophisticated manner. The principal signal currently used by nonmilitary receivers is the so-called L1 signal at 1575.42 MHz. This frequency is an integral multiple of 10.23 MHz, a fundamental frequency synthesized from an atomic clock aboard each satellite. The satellite's transmitter impresses upon this sinusoidal carrier wave a unique digital code sequence (the coarse-acquisition or C/A code), repeated once each millisecond.

The bits are encoded by reversing the phase of the carrier wave for a 1, and leaving the phase unchanged for a 0. This choice of encoding mode is important, because the phase of an electromagnetic wave is a relativistic scalar. The phase reversals correspond to physical points in spacetime at which--for all observers--the electric and magnetic fields vanish.

For the L1 signal, each bit lasts 1540 carrier cycles. This rather large number of cycles is not wasted; much of it is used for a very fast encrypted military code, 90° out of phase with the C/A code. Civilian navigation information is encoded on top of the C/A code at 50 bits per second. The navigation data include information from which the satellite's position and clock time can be accurately determined, and an almanac from which approximate positions of other GPS satellites can be computed. The timing signal corresponds to a phase reversal at a particular place in the navigation code sequence.

Every civilian GPS receiver carries circuitry that lets the receiver generate code sequences corresponding to the C/A code sequences from all the satellites. Many such sequences can be generated in parallel, depending on the sophistication of the receiver.

Because the satellite is moving with respect to the receiver, there is a first-order Doppler shift of the received carrier signal, of order v/c » 10^{-5}. A receiver may incorporate hundreds, or even thousands, of correlators that search in parallel through different frequency shifts and time offsets by comparing its own code sequence with those it receives. When an appropriately high correlation is found, the receiver locks onto the signal. The uniqueness of the transmitted code sequences lets the receiver identify which satellite a signal is coming from. First-order Doppler shifts, sometimes measured to within a few hertz, are used by some receivers to aid in extrapolating navigation solutions forward in time.

With the receiver locked onto a signal and the Doppler shift matched, timing information is obtained by comparing the receiver's clock time t_r with the time ticks encoded in the signal, thus measuring the "pseudoranges" $c(t_r - t_i)$, which are simply related to the right side of equations 1.

The relativistic fractional frequency shifts that concern us most--for example, the second-order Doppler shifts due to the motion of the orbiting clocks relative to the receivers--are a few parts in 10^{10}. These clocks are also very high up in Earth's gravity field and therefore suffer a gravitational frequency shift, given by

$$\frac{\Delta f}{f} = \frac{\Delta\Phi}{c^2}, \quad (4)$$

where DF is the gravitational potential difference between the satellite and the geoid. This gravitational shift causes clocks in GPS satellites to run faster than otherwise identical clocks on the ground by about 5×10^{-10}. Furthermore, because none of the orbits is perfectly circular, a satellite speeds up, or slows down, to conserve angular momentum as its distance from Earth varies along its orbit. That Keplerian variation periodically changes the second-order Doppler shift, while changing the gravitational frequency shift in the same sense.

Diurnal rotation and the Sagnac effect

Computations of satellite orbits, signal paths, and relativistic effects appear to be most convenient in an ECI frame. But navigation must generally be done relative to Earth's surface. So GPS navigation messages must allow users to compute satellite positions in an Earth-fixed, rotating coordinate system, the so-called WGS-84 reference frame.[6]

The navigation messages provide fictitious orbital elements from which a user can calculate the satellite's position in the rotating WGS-84 frame at the instant of its signal transmission. But this creates some subtle conceptual problems that must be carefully sorted out before the most accurate position determinations can be made. For example, the principle of the constancy of *c* cannot be applied in a rotating reference frame, where the paths of light rays are not straight; they spiral.

One of the most confusing relativistic effects--the Sagnac effect--appears in rotating reference frames.[7] (See Physics Today, October 1981, page 20*.) The Sagnac effect is the basis of the ring-laser gyroscopes now commonly used in aircraft navigation. In the GPS, the Sagnac effect can produce discrepancies amounting to hundreds of nanoseconds.

Observers in the nonrotating ECI inertial frame would not see a Sagnac effect. Instead, they would see that receivers are moving while a signal is propagating. Receivers at rest on Earth are moving quite rapidly (465 m/s at the equator) through the ECI frame. Correcting for the Sagnac effect in the Earth-fixed frame is equivalent to correcting for such receiver motion in the ECI frame. Suppose one sends a radio wave in a circle around the equator, from west to east, in an attempt to synchronize clocks along the path, invoking the constancy of *c*. Observers in the ECI frame see this wave propagating eastward a distance *x* in time *x*/*c*. Clocks in the signal's path move away from the wavefront with speed $W_E R$, where W_E is Earth's angular velocity and *R* is its radius. The distance such a clock moves in a time *x*/*c* is $W_E Rx/c$, and it takes an additional time $W_E Rx/c^2$ for the beam to catch up. For one complete circuit of the equator, this additional time is about 200 ns. Sending the signal in the opposite direction reverses the effect's sign.

The Sagnac effect also occurs if an atomic clock is moved slowly from one reference station on the ground to another. For slow clock transport, the effect can be viewed in the ECI frame as arising from a difference between the time dilation of the portable clock and that of a reference clock whose motion is solely due to Earth's rotation. Observers at rest on the ground, seeing these same asymmetric effects, attribute them instead to gravitomagnetic effects--that is to say, the warping of spacetime due to spacetime terms in the general-relativistic metric tensor. Such terms arise when one transforms the invariant ds^2 from a nonrotating reference frame to a rotating frame.[7]

Thus, attempts to establish a network of synchronized clocks on Earth's surface are subject to asymmetric, path-dependent effects arising from the planet's rotation. When atomic clocks became accurate enough for these effects to be significant, various proposals were made to deal with the Sagnac effect. One such proposal involved placing a discontinuity in TAI at the International Date Line. But such a scheme would not avoid path-dependent effects.

Synchronization should be an equivalence relation. To make it so, one could use the coordinate time. To achieve consistently synchronized clocks on Earth's surface at the subnanosecond level, the Consultative Committee for the Definition of the Second and the International Radio Consultative Committee have agreed that the correction term to be applied for the Sagnac effect should be $2W_E A_E/c^2$, where A_E is the projected area on Earth's equatorial plane swept out by a vector from Earth's center to the position of the portable clock or signal pulse.[3,7] (A_E is taken to be positive if the head of the vector moves eastward.)

The Sagnac effect is particularly important when GPS signals are used to compare times of primary-reference cesium clocks at national standards laboratories far from each other. Because their locations are very precisely known, each laboratory needs only one of the equations 1 to obtain GPS time from a satellite. The measurements are made in "common view"--that is to say, one satellite is observed simultaneously by receivers at two widely separated laboratories. When one takes time differences, many common-mode errors cancel out, yielding quite accurate time comparisons between remotely situated primary clocks. A Sagnac correction is needed to account for the diurnal motion of each receiver during signal propagation. In fact, one can use the GPS to observe the Sagnac effect.[8] Of course, if one works entirely in the nonrotating ECI frame, there is no Sagnac effect.

Gravitational and motional effects on clocks

Clocks in GPS satellites, being in different states of motion at various heights above the geoid, are subject to varying gravitational and time-dilation frequency shifts. To synchronize these clocks, one generates coordinate time *t*, as defined in the ECI frame by equation 3. Ignoring correction terms of order smaller than c^{-2}, one solves equation 3 for *dt* and integrates to get

$$\int_{\text{path}} dt = \frac{1}{c} \int_{\text{path}} \left[1 - \frac{\Phi - U}{c^2} + \frac{v^2}{2c^2} \right] ds. \quad (5)$$

Effects that contribute significantly to the integral on the right include Earth's gravitational potential (including quadrupole terms) and the satellite's velocity in the ECI frame.

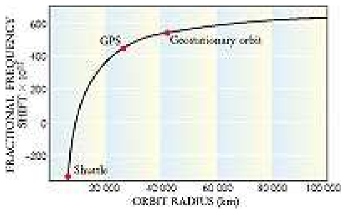

Figure 4

If the GPS orbits were perfectly circular, the corrections would include just a few constant contributions: for the gravitational potential differences between the satellites and the geoid, and for the second-order Doppler differences between the orbiting clocks and the reference clocks on the ground. Figure 4 shows how the relativistic frequency shift depends on the circular orbit's radius. At a radius of 9550 km, about 3000 km above the ground, the gravitational and Doppler effects cancel. Because the GPS orbits are higher than that, the gravitational blueshift is the largest contribution. So the net frequency correction for a GPS satellite is negative, amounting to 4.4645 parts per ten billion.

Nowadays the rate of every orbiting GPS clock is adjusted by this "factory offset" before launch. But before the first GPS satellite was launched in 1977, although it was recognized that orbiting clocks would require such a relativistic offset, there was uncertainty as to its magnitude, and even its sign. So correcting frequency synthesizers were built into the clocks, spanning a large enough range around the nominal 10.23 MHz clock frequency to encompass all possibilities. After the satellite's cesium atomic clock was turned on, it was operated for three weeks to measure its rate. The frequency shift measured during this initial period was found[9] to be 4.425 parts per ten billion, agreeing with the relativistic calculation to better than 1%.

Additional small frequency offsets arise from clock drift, environmental changes, and other unavoidable effects such as the inability to launch the satellite into an orbit with precisely the desired semimajor axis. The satellite clock frequencies are adjusted so that they remain as close as possible to the frequency of the Naval Observatory's clock ensemble. Because of such adjustment, it would now be difficult to use the GPS to measure the relativistic frequency shifts.

During the early days of GPS development, I found that the small diurnal-rotation contribution to the frequency offset had been inadvertently omitted. But eight years passed before system specifications were changed to reflect the correct calculation. The required change in the factory offset was a not-insignificant -1.2×10^{-12}.

If a clock's orbit is not perfectly circular, gravitational and motional frequency shifts combine to give rise to the so-called eccentricity effect, a periodic shift of the clock's rate with a period of almost 12 hours and an amplitude proportional to the orbit's eccentricity. For an eccentricity as small as 1%, these effects integrate to produce a periodic variation of amplitude 28 ns in the elapsed time recorded by the satellite clock. If it is not taken properly into account, the eccentricity effect could cause an unacceptable periodic navigational error of more than 8 m.

During the early development of the GPS, onboard computers had limited capability. It was decided that correcting for the eccentricity effect would be left to the receivers. Inexpensive receivers accurate to no better than 100 m may not need to correct for the eccentricity effect. But, for the best positional accuracy, receiver software must apply a relativistic eccentricity correction to the time signals broadcast by each satellite. The navigation message includes the current eccentricity, orbital elements, and other information that the receiver needs for computing and applying the eccentricity correction.

Rotating and inertial local frames

Information in the GPS navigation messages lets the user compute a satellite's position in the rotating WGS-84 frame at the moment of a signal's transmission. A typical receiver measures, on its own clock, the time differences $t_r - t_i$ for signals it receives from four or more GPS satellites. If these signals arrive simultaneously, the process is called receiver time tagging.

Generally, however, the transmissions arrive at different times. The navigation messages then let the receiver compute the position of each transmission event in the Earth-fixed WGS-84 frame. Before equations 1 can be solved to find the receiver's location, the satellite positions must be transformed to a common Earth-centered inertial frame, since light propagates in a straight line only in an inertial frame. These computations are performed by receiver software, producing relativistic corrections proportional to $1/c^2$.

When four or more GPS satellites simultaneously (according to *their* synchronized clocks) transmit signals to the same receiver, the procedure is called transmitter time tagging. The arrival times at the receiver may differ by as much as 18 ms, depending on the transmitter positions. The receiver must then keep track of its own motion during this receiving interval and make appropriate corrections. These corrections are again proportional to $1/c^2$, that is to say, they are also relativistic.

The strategies that receiver designers use to correct for these relative motions are not standardized. Choices depend, among other things, on the applications for which a particular receiver is intended. Some receivers correct signal arrival times for the Sagnac effect. That's most appropriate for receivers at fixed, well-measured locations. For example, GPS receivers at the NIST standards laboratory in Boulder, Colorado, serve to compare GPS time with time kept by NIST's ensemble of atomic clocks.

Confusion and consternation

Historically, there has been much confusion about properly accounting for relativistic effects. And it is almost impossible to discover how different manufacturers go about it! In one case, a manufacturer was found to be double-counting.[10] During 1989-90 I wrote letters to about a dozen receiver manufacturers inquiring about relativistic corrections in their software. Two of them responded with reasonable information, but nothing was heard from the others until some years later, when a rumor began circulating, alleging that some manufacturers thought I was trying to steal their secrets!

Another story, some years after that, had it that my letter caused consternation and much tweaking of receiver software. GPS managers have been extremely sensitive to assertions that relativistic effects were not being properly taken into account. Looking into these issues in 1985, the JASON group and a US Air Force Studies Board Committee found no significant omissions.

A 1995 meeting sponsored by the Army Research Laboratory considered the case of a rapidly moving GPS receiver. Did one, in such a case, need a coordinate system with its origin attached to the receiver in order to properly deal with clock synchronization? From the fast-moving receiver's point of view, it would seem that the GPS satellite clocks would not be synchronized. One can estimate the discrepancies from the approximate synchronization correction: vx/c^2, where v is the receiver's speed through the ECI frame and x is its distance from the GPS satellite in question. Suppose the receiver is itself in low Earth orbit (7.6 km/s) and the GPS transmitter is 20 000 km ahead. Then the synchronization correction comes to 1.7 ms. That's enough time for an electromagnetic signal to travel 500 m, so one would have to correct for it.

Within the framework of general relativity, however, one coordinate system should be as legitimate as another. Measurements made by an observer traveling with a moving receiver can just as well be described in another reference frame, by using transformations that relate the two frames. In the special case of two inertial frames in uniform relative motion, these are the familiar Lorentz transformations.

The TOPEX experiment

As one result of such considerations, William Feess of the Aerospace Corp proposed that data from the TOPEX/Poseidon satellite, launched into low-Earth orbit in 1992, could be used to test the relativistic predictions for the eccentricity effect on GPS clocks. TOPEX/Poseidon is a US-French mission to measure sea surface topography by radar altimetry. It carries an advanced six-channel GPS receiver capable of measuring both carrier phase and pseudorange for precise orbit determination. Additional, independent data for orbit calculation and tracking, as well as for ionospheric radar corrections, are provided by the DORIS Doppler tracking network.

The redundancy provided by six receiver channels lets one measure the eccentricity effect and check whether any other significant relativistic effects have been neglected. At least four channels are required for determining the TOPEX satellite's position and calibrating its clock. That clock can vary as much as 200 ns in a day as a result of instrumental drift and noise. So, to measure the eccentricity effect, one has to correct for this instrumental variation in order to generate the paper "coordinate clock" that keeps GPS coordinate time for TOPEX. Although this task requires only four GPS satellites, TOPEX is generally receiving data from six satellites. The additional data further constrain TOPEX's instantaneous position and clock time, and determine the elapsed proper time.

Figure 5 shows the results of such an experiment with data from the TOPEX/Poseidon receiver during 22 October 1995. The figure compares the predicted relativistic eccentricity effect for the GPS satellite that had the largest orbital eccentricity (almost 1.5%) on that day, with its measured time offset expressed as a distance. The data points are bunched because TOPEX, with its 90-minute orbital period, passes under the observed GPS satellite 11 times a day. All the other GPS satellites in the constellation showed the same good agreement with the relativistic prediction, but with smaller swing amplitudes corresponding to their smaller eccentricities.

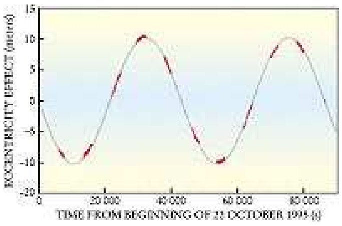

Figure 5

Marvin Epstein and coworkers at ITT have recently pointed out that the adjustments occasionally made to GPS orbits provide an opportunity for sensitive tests of small relativistic effects.[11] For example, rockets fired two years ago to reposition one of the GPS satellites reduced its semimajor axis by 1880 meters and increased its velocity correspondingly. Epstein and company observed an average fractional frequency change of -18.5×10^{-14}, in good agreement with the relativistic prediction of -17.7×10^{-14}.

Other small relativistic effects include the effect on satellite clock frequencies of the quadrupole moment of Earth's mass distribution, differences between coordinate distances and relativistically invariant distances, and the Shapiro time delay.[4,12] The Shapiro delay is the slowing of electromagnetic waves as they near Earth. For clocks in GPS orbit, this time delay is less than 200 ps. In the future, these very small effects will probably have to be incorporated into GPS calculations. Whether the calculations will be done by the system's master control station or delegated to a new generation of receivers remains to be seen.

Applications

The variety of GPS applications is astonishing. In addition to the more obvious civilian and military applications, the system's uses include synchronizing of power-line nodes to detect faults, very-large-baseline interferometry, monitoring of plate tectonics, navigation in deep space, time-stamping of financial transactions, and tests of fundamental physics. Two years ago, the value of the GPS to the general community had already become so great that President Bill Clinton turned off "selective availability"--the system by which the highest GPS precision was available only to the military.

At the Arecibo radio telescope in the 1970s and 1980s, Joseph Taylor and colleagues verified the general-relativistic prediction for the loss of energy by a binary pulsar through gravitational radiation. (See Physics Today, December 1993, page 17.*) Their exquisitely precise long-term timing measurements made use of the GPS to transfer time from the Naval Observatory and NIST to the local reference clock at Arecibo. The GPS constellation of highly stable clocks in rapid motion will doubtless provide new opportunities for tests of relativity. More than 50 manufacturers produce more than 350 different GPS products for commercial, private, and military use. More than 2 million receivers are manufactured each year. New applications are continually being invented.

Relativity issues are only a small--but essential--part of this extremely complex system. Numerous other issues must also be considered, including ionospheric and tropospheric delay effects, cycle slips, noise, multipath transmission, radiation pressure, orbit and attitude determination, and the possibility of malevolent interference.

Relativistic coordinate time is deeply embedded in the GPS. Millions of receivers have software that applies relativistic corrections. Orbiting GPS clocks have been modified to more closely realize coordinate time. Ordinary users of the GPS, though they may not need to be aware of it, have thus become dependent on Einstein's conception of space and time.

Neil Ashby *is a professor of physics at the University of Colorado in Boulder. Since 1974, he has been a consultant to NIST, Boulder, on relativistic effects on clocks.*

References

1. See the STEP homepage at http://www.sstd.rl.ac.uk/fundphys/step/.
2. *Navstar Analysis Update*, rep. no. 40-05, US Naval Research Laboratory, Washington, DC (June 2001).
3. N. Ashby, D. W. Allan, *Radio Sci.* **14**, 649 (1979).
4. N. Ashby, J. Spilker Jr, in *The Global Positioning System: Theory and Applications*, vol. 1., B. Parkinson, J. Spilker Jr, eds., American Institute of Aeronautics and Astronautics, Washington, DC (1997), p. 623.

5. N. Ashby, B. Bertotti, *Phys. Rev. D* **34**, 2246 (1986).
6. Defense Mapping Agency, *Department of Defense World Geodetic System 1984*, rep. no. DMA TR 8350.2, Defense Mapping Agency, Fairfax, Va. (September 1991). S. Malys, J. Slater, in *Proc. ION-GPS-94, Salt Lake City, 1994*, Institute of Navigation, Alexandria, Va. (1995).
7. L. Landau, E. Lifshitz, *The Classical Theory of Fields*, 4th ed., Pergamon, New York (1997), p. 254.
8. D. Allan, N. Ashby, M. Weiss, *Science* **228**, 69 (1985).
9. J. Buisson, R. L. Easton, T. B. McCaskill, in *Proc. 9th Annual Precise Time and Time Interval Applications and Planning Meeting*, L. Rueger et al., eds., Tech. Info. Div., NASA Goddard Space Flight Center, Greenbelt, Md. (1977), p. 177.
10. C. Alley, in *Proc. NATO Advanced Study Institute on Quantum Optics and Experimental General Relativity, August 1981, Bad Windsheim, Germany,* P. Meystre, M. O. Scully, eds., Plenum, New York (1983), p. 363.
11. M. Epstein, E. Stoll, J. Fine, in *Proc. 33rd Annual Precise Time and Time Interval Systems and Applications Meeting, Long Beach, Nov. 2001*, L. Breakiron, ed., US Naval Observatory, Washington, DC, (in press).
12. I. Shapiro, *Phys. Rev. Lett.* **13**, 789 (1964).

Related Readings from Physics Today

October 1981, page 20
December 1993, page 17
October 1999, pages 38
October 1999, page 44
January 1994, page 9
April 1993, page 9
March 2001, page 37