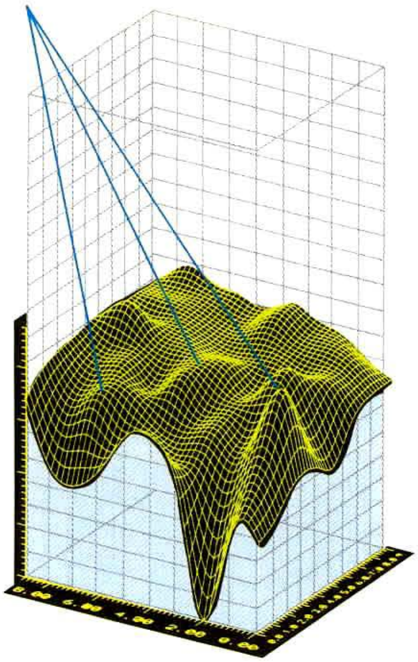

Time, Clocks, and GPS

Richard B. Langley

University of New Brunswick

"Innovation" is a regular column in GPS World *featuring discussions on recent advances in GPS technology and its applications as well as on the fundamentals of GPS positioning. In this issue we present a tutorial on one of the underlying principles of GPS: the precise measurement of time.*

This column is coordinated by Richard Langley and Alfred Kleusberg of the Department of Surveying Engineering at the University of New Brunswick. We welcome your comments and suggestions of topics for future columns.

Time is of the essence. This statement, used primarily to characterize the clauses in certain legal documents, also nicely describes the nature of the Global Positioning System. The basic measurement a GPS receiver makes is of the time required for a signal to propagate from a particular satellite to the receiver. Multiplying this time interval by the speed at which the signal propagates — the speed of light — converts it into a range or distance. Such a one-way ranging technique requires accurate timekeeping in both the satellite and the receiver. In this article, we'll investigate how this accuracy is achieved and examine some of the intricacies of time that are important in GPS positioning.

The high positioning accuracy of GPS is due, in part, to the use of atomic clocks to control the generation of the signals transmitted by the satellites. For redundancy, each Navstar GPS Block II satellite contains four atomic clocks, one of which is selected by the spacecraft controllers to provide the frequency and timing requirements for generating the satellite's signals. But just what is an atomic clock? Before we answer this question, let's examine some of the basic concepts associated with clocks and timekeeping.

THE QUARTZ CRYSTAL RESONATOR

All clocks contain an oscillator, which in turn contains a frequency-determining element called a resonator. A resonator is any device that vibrates or oscillates with a well-defined frequency when excited, such as guitar and violin strings, pendulums, and quartz crystals. Once excited, the oscillations of a resonator slowly die out due to energy loss. A perfect resonator has no energy loss and will oscillate forever.

To be useful in an oscillator, a resonator must be stimulated or energized repeatedly. This process is accomplished in a quartz crystal oscillator by taking advantage of the *piezoelectric effect*: An external voltage applied across opposite faces of a piece of quartz crystal cut in a prescribed way causes the crystal to expand or contract depending on the polarity of the voltage. The inverse is also true: forcibly deforming the crystal causes a small electrical potential to develop. A crystal connected to an alternating voltage source will vibrate. The vibrations, in turn, will generate an alternating voltage. These generated signals interact with the applied voltage in such a way that the vibrations and the resultant current flow are at a maximum at a particular frequency — the resonant frequency of the crystal, which is determined by the size of the crystal and how it is cut. Frequencies range from 10 kHz to well into the VHF range.

Accuracy (how well the oscillator can be tuned to a specified frequency) and stability (how well it stays on frequency) determine the quality of an oscillator. The oscillator's stability is measured in terms of the relative change in its frequency over a certain period of time. A high-quality quartz crystal oscillator kept at a constant temperature by a miniature oven has short-term stability (over periods shorter than about an hour) of a part in 10^{12} or better. But over the long term, its frequency can drift by several parts in 10^{11} per day. In order to keep two clocks using quartz crystal oscillators synchronized to 1 microsecond, you would have to reset them at least every few hours.

The resonators used in atomic clocks have surpassed considerably the accuracy and stability of quartz resonators.

ATOMIC RESONATORS

An atomic clock contains an oscillator whose oscillations are governed by a particular atomic process. According to the quantum picture of matter, atoms and molecules exist in well-defined energy states. An atom that falls from a higher to a lower energy state emits radiation in the form of light or radio waves with a frequency that is directly proportional to the change in energy of the atom. Conversely, an atom that jumps from a lower energy state to a higher one absorbs radiation of exactly the same frequency. The existence of such quantum jumps means that atoms can be used as resonators to govern clocks precisely, given a way to tap their resonance. Scientists found such a way in the late 1940s, using the ammonia molecule as a resonator. However, although the ammonia molecule was a very good resonator compared to what had been previously used in clocks, certain problems resulted in the resonant frequency being somewhat unstable. So scientists turned to the cesium atom and by the mid-1950s had built the first cesium clocks.

The cesium atom, in its naturally occurring form, Cs 133, consists of a nucleus containing 55 protons and 78 neutrons surrounded by a swarm of electrons. The outermost electron is in a shell of its own. Both the nucleus and this outer electron spin on their axes and, being charged, generate magnetic dipoles. In other words, they act like tiny bar magnets. The electron's magnetic dipole is either parallel to the nucleus's dipole (both north poles pointing in the same direction) or antiparallel. These two orientations correspond to two energy states of the atom, and transitions between these states form the basis of the cesium clock.

A block diagram of a cesium clock is shown in Figure 1. Cesium, a soft, silvery metal, is heated in an oven to a couple of hundred degrees Celsius. Individual atoms boil off and pass through a magnetic state selector, which deflects atoms in the higher energy state, allowing only atoms in the lower energy state to enter a microwave resonance cavity. While in this cavity, which operates on the same principle as the home micro-

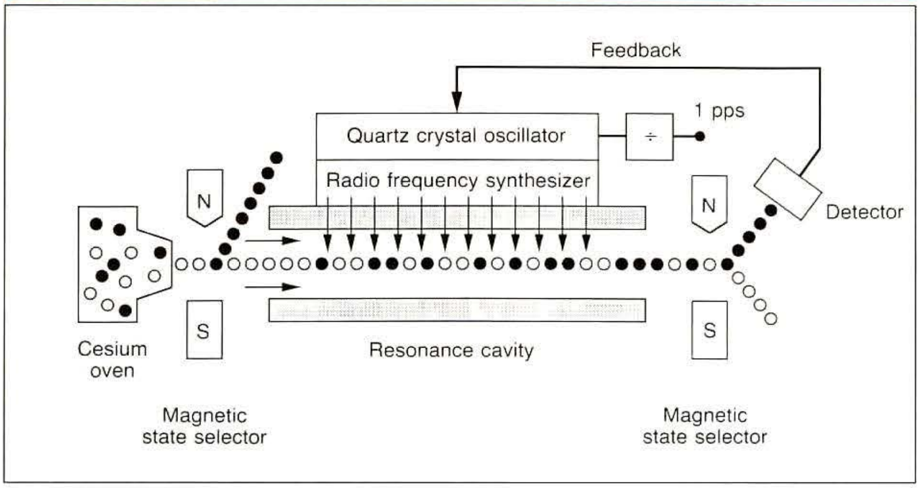

Figure 1. Basic operation of a cesium atomic clock

wave oven, the cesium atoms are subjected to a radio signal synthesized from the output of a quartz crystal oscillator. The frequency of this signal is very close to that of the energy difference between the two states of the cesium atom described above. A certain number of atoms absorb this radiation and change their state.

When the atoms leave the cavity, they again go through a magnetic gate, which directs atoms in the higher energy state toward a detector. The detector produces an electrical signal whose intensity is related to the number of atoms it intercepts. This signal is fed back to the quartz crystal oscillator and is used to control the oscillator's frequency and, hence, that of the synthesized radio signal. Through this feedback process, the frequency is automatically adjusted to maximize the number of atoms reaching the detector, which means that the radio frequency exactly equals the cesium atom's resonance frequency. The frequency of the oscillator's output signal can be electronically divided down to produce, for example, a one-pulse-per-second signal to drive a display that tells the time.

Each GPS Block II satellite contains two cesium clocks, with typical stabilities of 1 to 2 parts in 10^{13} over a one-day period. These clock stabilities contribute from 2.6 to 5.2 meters to the pseudorange error budget. Over an averaging period of 10 days, the stability of the cesium clocks improves to about 4 parts in 10^{14}, decreasing only slightly for periods as long as 100 days or more. Cesium clocks are well known for their excellent long-term stability.

Atomic clocks also have been developed based on two other resonators. The rubidium clock is based on a particular resonant frequency of rubidium atoms. The atoms, in the form of a very low pressure gas, are contained in a glass cell situated inside a microwave resonance cavity. A microwave signal whose frequency is tuned by a feedback loop induces transitions. The short-term stability of rubidium clocks is almost as good as that of cesium clocks. In fact, over one-day averaging periods, the rubidium clocks in some of the Block I satellites had stabilities of 1.6 parts in 10^{13} or better. However, the frequency of rubidium clocks tends to wander over longer periods, resulting in poorer performance. Each Block II satellite contains two rubidium clocks in addition to the two cesium clocks.

The third and most stable of the atomic clocks is the hydrogen maser. The word *maser* is an acronym for microwave amplification by stimulated emission of radiation, and, as the name suggests, its operating principle is similar to that of the laser. According to the rules of quantum mechanics, an atom in an energy-emitting state will, eventually, emit radiation of its own accord. Hydrogen masers magnetically select hydrogen atoms in an energy-emitting state to enter a quartz storage bulb. If enough atoms are present in the bulb, one of them will emit spontaneously a packet of radiation — a photon — at the resonant frequency. A photon that strikes another atom in the energy-emitting state may stimulate that atom to emit radiation at exactly the same frequency and exactly in phase with the incident radiation. This process continues with other atoms, and self-oscillation starts very rapidly in the resonance cavity surrounding the bulb. A detector picks up the resulting microwave signal and uses it to phase lock a crystal oscillator.

Hydrogen masers have been built with stabilities on the order of a few parts in 10^{15}. However, masers are not as common as cesium or rubidium clocks because they generally are more costly and less rugged. A maser has been employed in a sub-orbital rocket flight, but one has yet to be flown in a satellite. Plans to modify the Block II satellite qualification model for an Advanced Clock/Ranging Experiment include a hydrogen maser clock as one of the four clocks in the satellite.

JUST A SECOND

Before 1956, the fundamental unit of time was the *mean solar day*, which is based on the rotation of the earth on its axis. But it was known for some time that the earth's rotation speed, and hence the length of the day, was not constant due to the action of ocean tides, atmospheric winds, and even motions in the earth's core. Although the larger variations could be measured and time scales corrected to account for them, it became necessary to define a new fundamental unit of time. In 1956, the second, defined in terms of the period of the earth's orbit around the sun, became the new standard. But this standard, known as the *ephemeris second*, was short-lived.

Because the frequency of the cesium resonator is so stable, the development of the cesium clock made it possible to define the second such that it could be accurately realized in a laboratory without resorting to a long series of astronomical observations. In 1967,

Cesium clocks are well known for their excellent long-term stability.

the *atomic second* formally replaced the ephemeris second as the fundamental unit of time in the International System of Scientific Units. The atomic second was defined as exactly 9,192,631,770 cycles of the unperturbed microwave transition between the two energy levels of Cesium 133 used in the cesium clock, a number chosen to agree closely with the ephemeris second over the period 1956 to 1965 to avoid discontinuities in astronomical records.

The time scale based on the atomic second is known as International Atomic Time or TAI, the French acronym. However, not even an atomic clock keeps perfect time. So, rather than rely on just one clock to provide atomic time, TAI is computed from an ensemble of atomic clocks throughout the world. An international body called the Bureau International des Poids et Mesures (BIPM) based in Sèvres just outside Paris performs this computation. Through a variety of clock-comparison techniques, including Loran-C radionavigation signals, television signals, and GPS, the Bureau compares the readings of atomic clocks from more than 50 laboratories around the world. The definitive TAI is determined from those readings.

Not even an atomic clock keeps perfect time.

As an aside, we note that since 1983 the meter also is defined in terms of the atomic second. By fixing the speed of light at 299,792,458 meters per second, the 17th General Conference on Weights and Measures defined the meter as the distance travelled by light in a vacuum during 1/299,792,458 of a second, or a little over 3 nanoseconds.

UNIVERSAL TIME

Although an imperfect timekeeper, the earth's rotation with respect to the sun has governed human affairs since time immemorial. *True solar time*, the time kept by a sundial, is nonuniform due to the earth's varying speed in its orbit and because the earth's equator is not parallel but inclined to its orbit by about 23.5°. However, a mean time scale can be derived by calculating the variations in true solar time, which can amount to as much as 16 minutes, and removing them from true time. The origin of such a time scale, that is, the instant when the time is zero hours (midnight) or twelve hours (noon), is arbitrary. In fact, the custom up until the late 1800s was for each village, town, or city to define its own local mean time. Noon was when the "mean" sun crossed the local meridian. But with the introduction of standard time zones, everyone in a zone approximately 15 degrees of longitude wide kept the same time, an integral number of hours different from time at the Greenwich meridian — Greenwich Mean Time (GMT) or Universal Time (UT), as it has come to be called. However, a few exceptions to this rule exist. For example, standard time in the Canadian province of Newfoundland is three and a half hours behind UT.

Although corrected for the nonuniformity of the sun's apparent motion, mean solar time, and hence UT, is still irregular due to variations in the earth's spin on its axis, as mentioned earlier. A slight shifting of the rotation axis with respect to the earth's crust, a phenomenon known as *polar motion*, causes further complication. An observatory's determination of Universal Time from measurements of the passage of stars across its meridian, known as UT0 (UT zero), embodies the effects of both polar motion and variations in the earth's spin. Polar motion's contribution to UT0 is a function of the observatory's position on the earth, and only if a number of observatories around the world each determine UT0 can the magnitude of polar motion be determined and the UT0 measurements corrected. UT0 corrected for the effect of polar motion is known as UT1. UT1 represents the actual orientation of the earth in space and is the time needed, for example, by sailors and others to navigate by the stars.

But UT1 is still a nonuniform time scale. In addition to irregular short-term and seasonal variations, UT1 drifts with respect to atomic time in the long term. At present this drift amounts to several milliseconds per day. Over the span of a year these milliseconds can add up to a full second. Civil timekeeping required a time scale that had the uniformity of atomic time but was not too different from UT1. Such a time scale, called Coordinated Universal Time or UTC, was introduced in 1961. Originally UTC was kept close to UT1 by periodically adjusting its rate and by adding or subtracting steps of a fraction of a second. However, since 1972 the UTC rate has been set equal to that of TAI and leap seconds have been introduced into the UTC time scale to prevent UTC from deviating from UT1 by more than 0.9 seconds. The last leap second (as of this writing) was inserted into UTC just before midnight on the last day of December 1990, which then had 86,401 seconds. The insertion of that leap second put UTC exactly 26 seconds behind TAI. The provision also exists for removing leap seconds from UTC should UT1's rate of change reverse sign, as it did in the last century.

UT1 traditionally was derived from visual and photographic observations of the passage of stars across an observatory's meridian, with responsibility for definitively determining UT1 given, by international agreement, to the Bureau International de l'Heure (BIH) at the Paris Observatory. As the years passed, the BIH refined the reduction process and incorporated observations from new techniques that became available. Eventually, the more accurate space geodetic techniques of very long baseline interferometry and satellite and lunar laser ranging supplanted conventional techniques. In 1987, the activities of the BIH were reorganized, and the new International Earth Rotation Service (IERS), whose Central Bureau remained at the Paris Observatory, took over responsibility for determining UT1 and polar motion in 1988. The IERS works closely with the BIPM to determine when UTC leap seconds will occur.

GPS TIME

The signals transmitted by GPS satellites are referenced to *GPS* (System) *Time*, which until June 1990 was the time kept by a single atomic clock at one of the monitor stations. However, GPS Time now derives from a composite or "paper" clock consisting of all operational monitor station and satellite clocks.

GPS Time is steered over the long run to keep it within about 1 microsecond of UTC, ignoring leap seconds. So unlike UTC, GPS Time has no leap second jumps. At the integer second level, GPS Time equalled UTC in 1980, but presently, due to the leap seconds that have been inserted into UTC, it is ahead of UTC by 7 seconds plus a fraction of a microsecond that varies day to day.

A particular epoch is identified in GPS Time as the number of seconds that have elapsed since the previous Saturday/Sunday midnight. Such a time measure is, of course, ambiguous, so one must also indicate in which week the epoch is. GPS weeks start with week 0 on January 6, 1980, and are numbered consecutively. Figure 2 (a screen shot of a Macintosh Computer HyperCard stack) shows the GPS week and seconds of the week corresponding to the UTC epoch of 12:00:00, November 1, 1991; the corresponding Julian Date (JD), which represents the number of days and fractional days elapsed since noon UT on January 1, 4713 B.C.; and the Modified Julian Date (MJD), which equals JD minus 2400000.5. JD and MJD are frequently used by astronomers, navigators, and others for compactly and unambiguously identifying a particular epoch in time.

Atomic clocks perform best with a minimum of adjustments. So to avoid continuous adjustment, the clocks in the GPS satellites are only approximately synchronized to GPS Time. The GPS Operational Control System and the United States Naval Observatory (USNO) carefully monitor the offsets of the satellite clocks from GPS Time, which can be a millisecond or so. They then determine an offset at an initial epoch, a linear drift term, and, for rubidium clocks, a drift rate of change term for each satellite clock. These parameters are uploaded to the corresponding satellite and subsequently included in its navigation message. A GPS receiver uses the satellite clock data to convert the measured pseudoranges from the satellite time scale to GPS Time. The satellite message also includes the offset of GPS Time with respect to UTC.

When a GPS receiver initially acquires signals, its clock, in general, will have a large unknown offset with respect to GPS Time. This offset, however, will contribute the same timing bias to all pseudorange measurements made at any particular epoch and can be solved for along with the receiver coordinates. Once determined, the bias can be used to synchronize the receiver clock to GPS Time. GPS Time or UTC then can be displayed by the receiver or used to time-tag recorded data or to generate a one-pulse-per-second electrical signal for controlling other equipment. Figure 3 shows schematically the relationships among the satellite, receiver, and system time scales and pseudorange measurements. The receiver makes a raw measurement of the time interval, $d\tau$, which when multiplied by the speed of light, c, gives the measured pseudorange, p. Correcting this measurement for satellite clock, dt, and receiver clock, dT, offsets with respect to GPS Time gives the true geometric range, ρ, ignoring propagation delays and other potential biases.

Several manufacturers offer GPS receivers specifically designed for use as sources of precise time information. Such receivers generally are operated from fixed sites and, once their locations are accurately determined, provide synchronized time signals even when only one satellite is in view. These receivers readily achieve accuracies to within about 100 nanoseconds.

The USNO, BIPM, and others have developed sophisticated techniques for using GPS to synchronize clocks to a precision of 10 nanoseconds or better even when the clocks are on different continents.

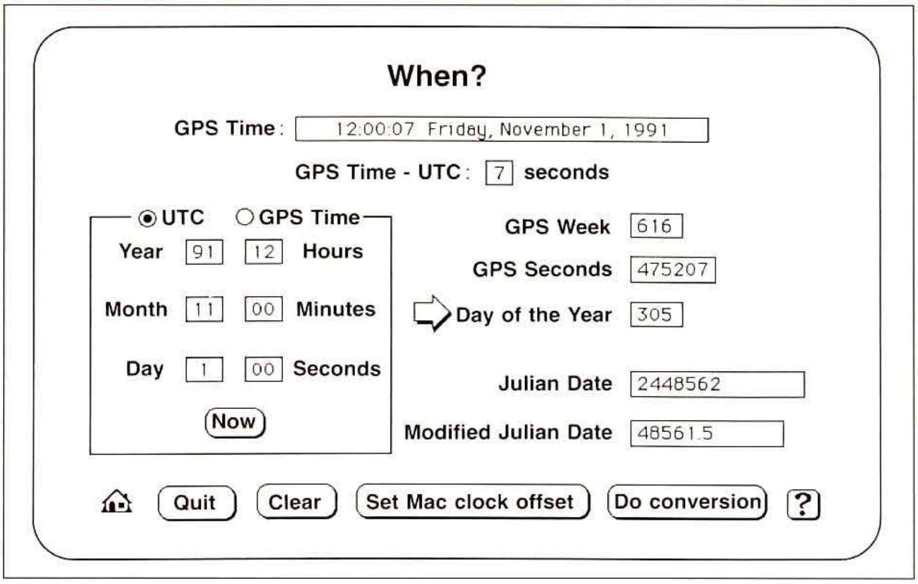

Figure 2. Computer screen display illustrating the conversion of an arbitrary UTC epoch to GPS Time

RELATIVISTIC EFFECTS

The atomic clocks in the GPS satellites have a 10.23-MHz frequency output. This fundamental frequency corresponds to the chipping rate of the pseudorandom noise P-code and, when divided by 10, gives the rate of the C/A-code. Multiplying the fundamental frequency by 154 produces the L1 carrier frequency and by 120 produces the L2 carrier frequency. Actually, in order to account for the effects of relativity, the fundamental frequency of the satellite clocks is set at slightly less than 10.23 MHz.

Einstein showed in his Special Theory of Relativity published in 1905 that a clock moving with a constant speed relative to another clock will appear to run more slowly. Accordingly, a clock in a satellite traveling in a circular orbit around the earth would appear to lose time compared to one on the ground. But 11 years later in his General Theory of Relativity, Einstein deduced that clocks in different gravitational potentials also will appear to run at different rates. Due to the difference in gravitational potential, a clock in a satellite will appear to run faster than one on the ground. The net effect on a satellite clock is the combination of the two effects. A clock in a GPS satellite in a circular orbit with a nominal radius of 26,560 kilometers gains 38.4 microseconds per day compared to one on the ground. This time difference corresponds to a relative frequency offset of its oscillator of 4.45×10^{-10}. In order to compensate for this offset, the fundamental frequency of the satellite clocks is reduced by 0.00455 Hz to 10.22999999545 MHz.

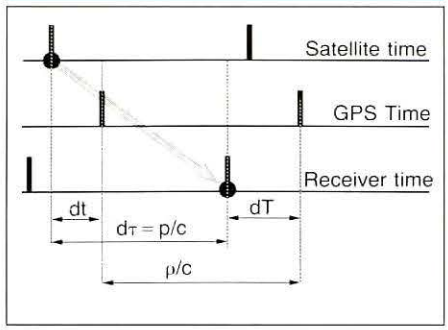

Figure 3. Relationships among GPS Time and satellite and receiver time scales

If GPS satellites were in circular orbits, their signals would require no further compensations for relativity to achieve ranging accuracies at the meter level. However, the orbital eccentricity of a GPS satellite can range up to 0.02, which means that both its speed and the gravitational potential it experiences change with time. The result is an oscillating time offset that is proportional to eccentricity and varies sinusoidally with the position of the satellite in its orbit. The magnitude of this effect can range up to 45.8 nanoseconds, which corresponds to a ranging error of 13.7 meters. A GPS receiver must correct its measured pseudoranges for this

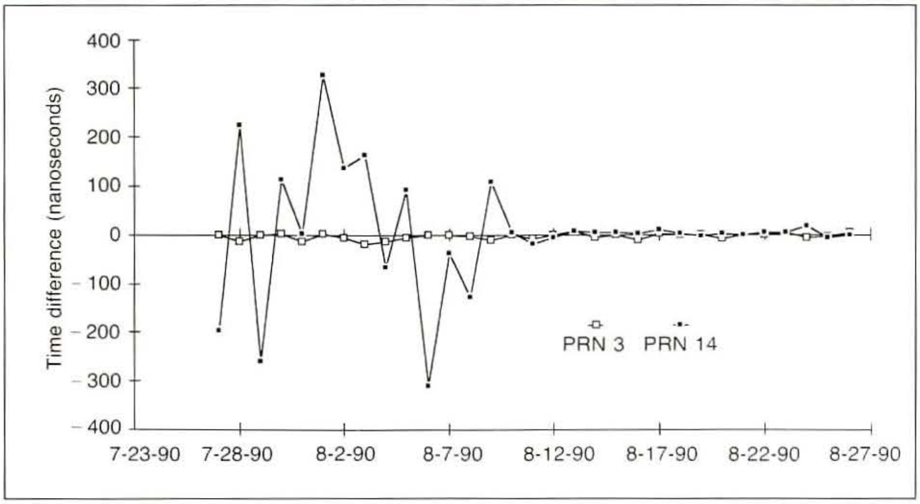

Figure 4. Discrepancies remaining in individual daily determinations of GPS Time at the Paris Observatory after removal of a smoothed estimate based on data from all Block I satellites

variation using the satellite orbit description contained in the navigation message.

SELECTIVE AVAILABILITY

The normally excellent behavior of atomic clocks in GPS satellites is intentionally degraded when selective availability (SA) is in effect. SA is one of the methods used by the Department of Defense (DoD) to deny the full accuracy capability of GPS to "nonauthorized users," meaning most civilians. In addition to purposely degrading satellite orbit information in the navigation message, DoD manipulates or *dithers* the satellite clock frequency. This dithering, referred to as the δ-process, introduces errors in the measured pseudoranges and carrier-phase measurements. Together with orbit errors, dithering can limit the accuracy of horizontal positions by up to 100 meters, 95 percent of the time. Excursions up to 300 meters may be expected the remaining 5 percent of the time. Velocity accuracy is also degraded.

SA also compromises the accuracy with which GPS Time or UTC can be determined from pseudorange measurements. Figure 4 shows the effect of SA on timing measurements made at the Paris Observatory in July and August 1990. The differences between daily determinations of GPS Time from two individual satellites (PRN 3, a Block I satellite; and PRN 14, a Block II satellite) and a best determination of GPS Time obtained by smoothing the data from all Block I satellites are plotted. The turning off of SA on about August 10, 1990, can be seen clearly.

Authorized users of GPS can recover from SA errors using information encrypted in the navigation message. Nonauthorized users must either endure SA or find ways of minimizing its effect. For example, by using differential GPS techniques, in which a pair of receivers simultaneously observe the same set of satellites, the effect of clock dithering can be differenced away (see "The Issue of Selective Availability," *GPS World*, September/October 1990).

CONCLUSION

In this article we have looked at the very important role that time plays in the Global Positioning System. Thanks to minute energy changes in individual atoms of cesium and rubidium, humankind possesses the ability to synchronize clocks anywhere in the world to better than 10 nanoseconds. But given this amazing ability to measure time, we still don't know what time actually is. What St. Augustine said at the end of the fourth century still holds: "What, then, is time? If no one asks me, I know what it is. If I wish to explain it to him who asks, I do not know." ■