HAROLD D. BLACK

SATELLITES FOR EARTH SURVEYING AND OCEAN NAVIGATING

The Navy Navigation Satellite System (Transit) pioneered the use of near-earth satellites for land survey and ship navigation. Currently, there are about 16,000 ships using the system for navigation.

Surely, and at an increasing pace, satellites are having important effects on our daily lives: the decreasing costs of long-distance telephone calls via communication satellites, the satellite-produced weather map on the evening news, and the Landsat-produced photographs of the earth are three examples. It is less well known that a satellite system has resulted in a dramatic improvement in the ancient arts of navigation and surveying.

The traditional forms of celestial navigation are limited by being useful only when the sun or stars are visible. Existing radio techniques are either unavailable or have questionable accuracy over certain portions of the earth.

Surveying on a global scale had been hampered by more subtle limitations—national boundaries, difficult terrain, and broad oceans—which thwarted attempts to place terrestrial points in a common reference frame. Even without those limitations, there were unresolved inconsistencies associated with the inability to make vertical measurements relative to an earth-centered coordinate system.

An ingenious technique for determining the path (orbit) of a satellite has provided an elegant solution to those problems and has, in turn, created a new industry: satellite navigation and surveying, whose total sales this year will exceed $60,000,000. The 20 year development of the technique has still not exhausted its potential accuracy. Development will probably continue for several more decades.

THE RUSSIAN CONTRIBUTION

In October 1957, shortly after the launch of Sputnik I, W. H. Guier and G. C. Weiffenbach, physicists at the Applied Physics Laboratory, tuned a communications receiver to the Sputnik frequency, approximately 20 MHz.[1] What they heard was a characteristic drop in pitch as the satellite approached and then receded into the distance. This "Doppler shift" in frequency[2] is familiar to anyone who has waited at a railroad crossing and heard the whistle as the train approached and then disappeared down the track. Other examples are the whistle of a low-flying jet and the falling pitch of a racing car as it whizzes by.

A characteristic Doppler curve is shown in Fig. 1. After an intensive effort, Guier and Weiffenbach showed that the shape of this curve, when used together with the laws of motion, contains all the necessary information to derive the satellite orbit. The Doppler shift was a familiar phenomenon to them. But when they tried to match the shape of the curve with successively more accurate (theoretical) descriptions of the satellite orbit, the data denied success until a rigorously correct description of the satellite motion was utilized. Then there was a match between the key and lock; there was only one orbit that would match the data.

But it wasn't as simple as this sounds. Guier and Weiffenbach made a number of original discoveries and innovations. For example, they recognized that

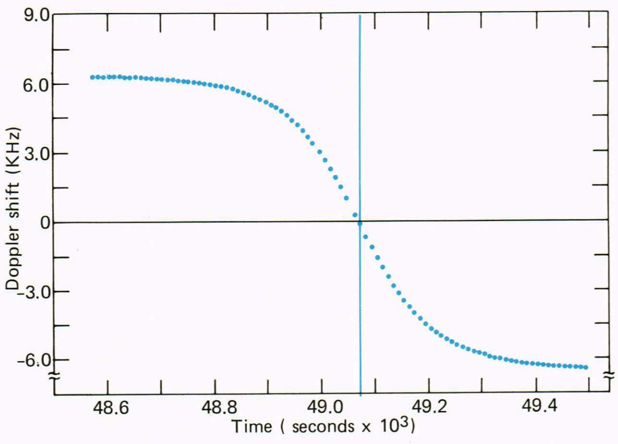

Fig. 1—Doppler shift of a 300 MHz satellite oscillator in an 1100 km orbit. The observing site is at 45° latitude, and the satellite is 77° above the observer's horizon at the time of closest approach. Zero Doppler shift occurs at that time.

meaningful frequency measurements would require that the satellite-borne oscillator and the ground oscillator (WWV) had to be calibrated against a common frequency standard. Of course that was impossible because both changed slowly with time and the satellite oscillator was hundreds of miles in space. They then discovered that they could perform the calibration with the satellite in orbit by utilizing the entire Doppler curve and by including the calibration as part of the orbit determination.[3]

On data obtained from Sputnik II, which was launched about a month later, they were able by means of a clever analysis to take advantage of two frequencies broadcast by the satellite (20 and 40 MHz) to remove an error caused by the ionosphere.

The navigation system was effectively invented when the late F. T. McClure, Chairman of the APL Research Center, realized that the discovery of Guier and Weiffenbach could be inverted, that is, solved in the reverse order. If they could at a *known* site derive the orbit of a satellite from Doppler shift measurements, then, using that orbit, navigators could at *unknown* sites derive coordinates from the orbit and Doppler shift measurements. The system should have some striking advantages over other forms of navigation:

1. Since the measurement of angles or directions are not required, simple nondirectional receiving antennas suffice. Directional antennas aboard a rolling, pitching ship are complicated and create a serious maintenance problem.
2. Since optical measurements are not involved, the system would be immune to the vagaries of the weather. For months on end, the skies over the northern Pacific and Atlantic Oceans are cloud covered. During such periods, celestial navigation is useless.
3. All the equipment sites that are required to operate the system could be within the U.S.A. This avoids the political and logistic problems associated with operating stations in foreign countries.
4. On land, repeated Doppler "navigations" at a fixed site become a new form of surveying. The earth could be globally surveyed in an internally consistent coordinate system. R. B. Kershner, the system architect, recognized the surveying possibilities inherent in Doppler navigation and named the system "Transit."

Guier and Weiffenbach describe the events surrounding their original discovery in this issue.

HOW THE SYSTEM WORKS

Figure 2 shows schematically the overall architecture and the four basic elements of the system.

A Constellation of Satellites

A number of satellites, five at present, are in near-earth orbits (Fig. 3) that pass over the earth's pole, are circular, and have an altitude of approximately 1100 km. Each satellite contains:

1. A highly precise frequency standard that drives two transmitters, nominally at 150 and 400 MHz. A counter driven by this same standard functions as a satellite clock.
2. A core memory that holds a current "ephemeris" of the satellite. An ephemeris is a table of satellite positions versus time. The ephemeris information and a clock control register can be revised from the ground by means of a communication channel. The satellite is oriented so that the antennas always face the earth. The ephemeris information is relayed to the navigator via modulation patterns on the 150 and 400 MHz transmissions, which are never turned off.

Tracking Stations

Four stations (Hawaii, California, Minnesota, and Maine) track the satellite signals at every opportunity. By "track" we mean that the stations measure the frequency of the satellite signal at four second intervals. After the satellite has set (typically, 17 minutes

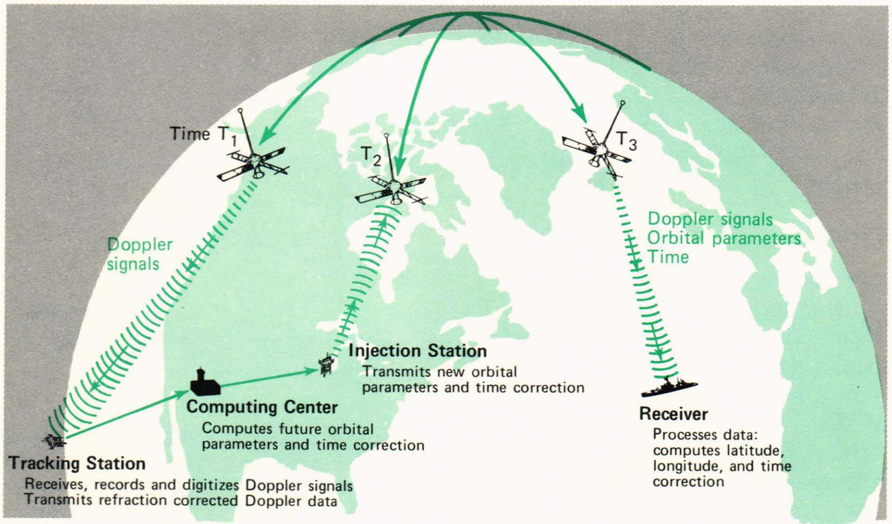

Fig. 2—System architecture of the Navy Navigation Satellite System (Transit).

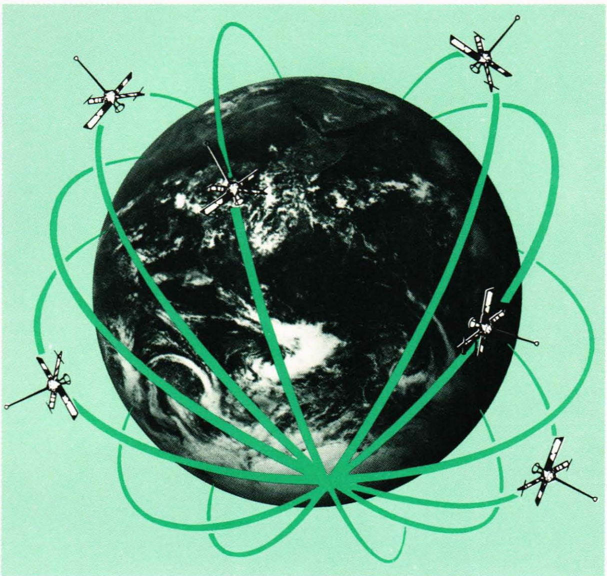

Fig. 3—Transit satellites in polar orbits at approximately 1100 km. The satellite period at this altitude is about 107 minutes. While six orbits are shown, there are currently (1981) five satellites in the constellation.

elapse from rising to setting), the measurements are transmitted to a central computing facility where all measurements for each satellite from the four tracking stations are accumulated. At least once a day they are used in a large computing program to:

1. Determine a contemporary orbit specification for the satellite and prepare an ephemeris for the next 24 hours.
2. Compute the necessary satellite clock corrections to compensate for the predictable part of the oscillator drift — typically, several parts in 10^{11} per day.
3. Calibrate all tracking station oscillators and clocks relative to a common standard.

The ephemeris prediction and satellite clock correction information are then transmitted back to one of the three injection sites.

The Injection Station

Each new ephemeris is inserted into the satellite memory by means of the radio command link, each injection writing over the one that is about to expire. (One station actually performs the injection while a second provides backup in case of equipment failure.) Injections are at 12-hour intervals since every satellite is visible at every station at least once every 12 hours. (The satellite memory has sufficient storage to contain a 16-hour ephemeris.)

The User

A surveyor/navigator measures the received frequency at discrete intervals and receives the satellite ephemeris broadcast by the satellite. With his frequency measurements, the ephemeris, and his own motion, he can compute his position. While the computation is neither simple nor amenable to hand computation, it is easily programmed for a small digital computer. We will describe the computation in a later section.

For the sake of brevity, we have omitted a number of details that can be found in the references.

DOPPLER POSITIONING

There are two kinds of position determination: navigation and surveying. These differ in several respects. Navigation is near-real-time positioning, usually only in two dimensions, aboard a moving ship. The navigator is in a hurry to get his position and is not concerned about errors that are smaller than the length of his ship. Surveying is establishing the position of an earth-fixed point, usually in three dimensions. Accuracy requirements for the latter are typically two orders of magnitude tighter than those of the former. Surveyors fix an antenna location and collect many passes of data for analysis, carefully balancing the selection of data to remove correlated errors (for example, with equal numbers of north-going and south-going passes).

FLAT-EARTH APPROXIMATION

If an observer precisely measures the frequency-time characteristic of a train whistle, the function shown on the billboard in Fig. 4, then with a timetable for the train (the train's "ephemeris"), the observer's position can be determined. This is a "flat earth" analog of satellite navigation. Because it correctly illustrates just how the positional information is contained in the Doppler shift — without bringing in the complexities of three-dimensional satellite motion — we will explain it in some detail. We seek the navigator's position in the coordinate system shown in Fig. 4. In our illustrative problem, we take the track to be straight, and the whistle frequency, f_T, as measured by an observer riding on the train, to be constant. Moreover, we are given a position-timetable for the train, in the (x-y) coordinate system shown in Fig. 4.

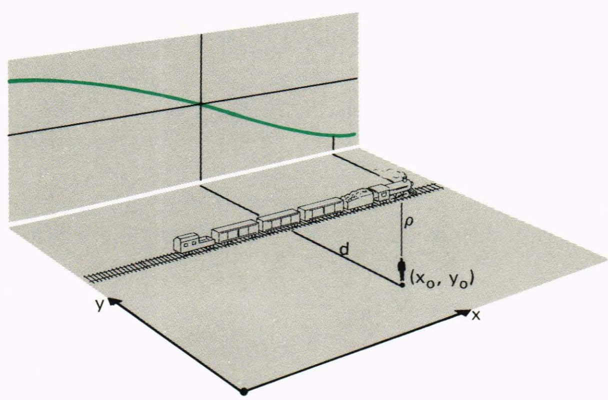

Fig. 4—The Doppler shift from a train whistle. At either end of the curve, it is asymptotic to $\pm f/c\,\dot{x}$, where f is the train whistle pitch, c is the speed of sound, and $\dot{x}$ is the train speed.

If we imagine a train (satellite) passing, the distance from the observer to the train (satellite) appears as in Fig. 5a. The range, ρ, decreases as the train (satellite) approaches, reaches a minimum, and then increases again. The negative slope of this function is shown in Fig. 5b. The Doppler shift imposed on the train whistle (satellite oscillator) is proportional to the function shown in Fig. 5b:

$$\Delta f = -\frac{f_T}{c}\frac{d\rho}{dt}, \tag{1}$$

where c is the phase "velocity" of the signal. (For the train whistle [sonic energy], c is the speed of sound; for the satellite [electromagnetic energy], c is the speed of light.) Based on the railroad track experience, the minus sign is easy to rationalize. As the train goes away from an observer ($d\rho/dt > 0$) and from observation, the frequency clearly decreases. The minus sign in Eq. 1 assures that the equation is consistent with this observation. Newton[1] gives an explanation of the Doppler shift on a more basic level of understanding.

The measured frequency is

$$f = f_T + \Delta f = f_T\left(1 - \frac{1}{c}\frac{d\rho}{dt}\right). \tag{2}$$

With f_T known, we can separate the Doppler shift,

$$\Delta f = -\frac{f_T}{c}\frac{d\rho}{dt}, \tag{3}$$

from the measurement. We consider the slope at the steepest part of this curve, the point at which $\Delta f = 0$ (see Fig. 5b); the rate of change of Δf is

$$\frac{d}{dt}(\Delta f) = -\frac{f_T}{c}\frac{d^2\rho}{dt^2}. \tag{4}$$

Let x be the along-track distance and d be the unknown distance from the track. We may write for ρ:

$$\rho = \sqrt{(x-x_0)^2 + (y-y_0)^2}, \tag{5}$$

and obtain an expression for $d^2\rho/dt^2$ by twice differentiating with respect to time. At the point where we have measured the slope of the Doppler curve, this expression simplifies to:

$$\frac{d^2\rho}{dt^2} = \frac{v^2}{\rho}, \tag{6}$$

in which $v \triangleq dx/dt$ is the speed of the train. Also at this point we have $\rho = d$; so, substituting from Eq. 6 to Eq. 4, we can solve for d:

$$d = -\frac{v^2 f_T}{c\left(\frac{d}{dt}\Delta f\right)}. \tag{7}$$

Thus, the observer's distance from the track is determined if we know the speed of the train (from the train's ephemeris) and the slope of the Doppler shift

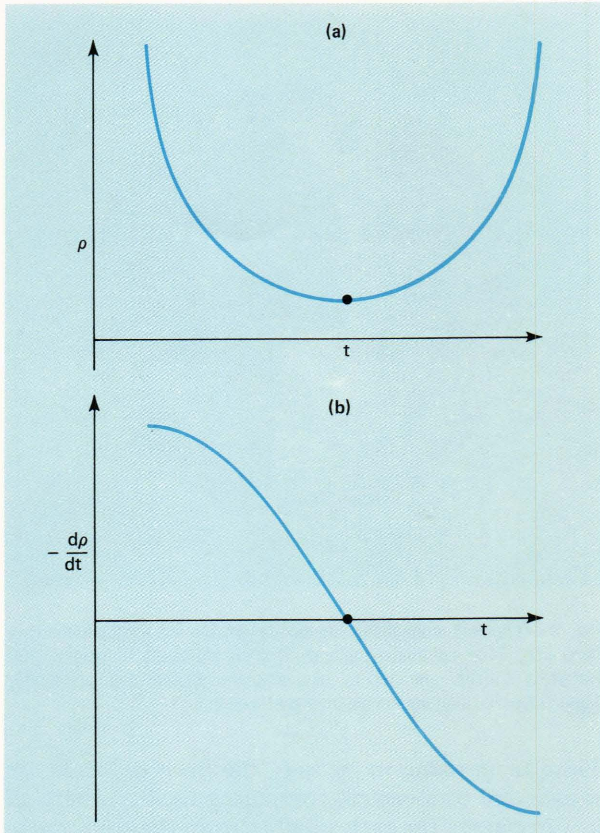

Fig. 5—Slant range versus time, shown in (a). The negative slope of the slant range is shown in (b). The Doppler shift is directly proportional to this latter quantity.

(from measurement) at the point where the shift is zero. The observer's y coordinate is simply $y_0 = y - d$. The x or along-track position component is easy to obtain. If we identify the time at which $\Delta f = 0$ and simply read the x-component of the train at that time, this is the x-component of the observer.

Like all analogies, this one has its limitations. Because of symmetry, we cannot resolve which side of the track the observer is on. In the satellite case, the earth's rotation introduces an asymmetry into the measurement and resolves the ambiguity. We have given the analogy here to illustrate that two coordinates of position are readily available from Doppler measurements. We could push the analogy a little further to illustrate that we do not need to know precisely the unshifted frequency of the whistle. If we measure the entire Doppler function and if the average of these measurements is not zero, then this nonzero value is a correction to our estimated whistle frequency. Additionally, because this is a flat-earth approximation, it fails to demonstrate that the third coordinate, observer height, can be determined in the real (satellite) case. This will be clarified in a later section.

The navigator does not utilize this simple-minded technique, but rather computes a fix from the entire Doppler shift curve and a least-squares fitting algorithm (Fig. 6).

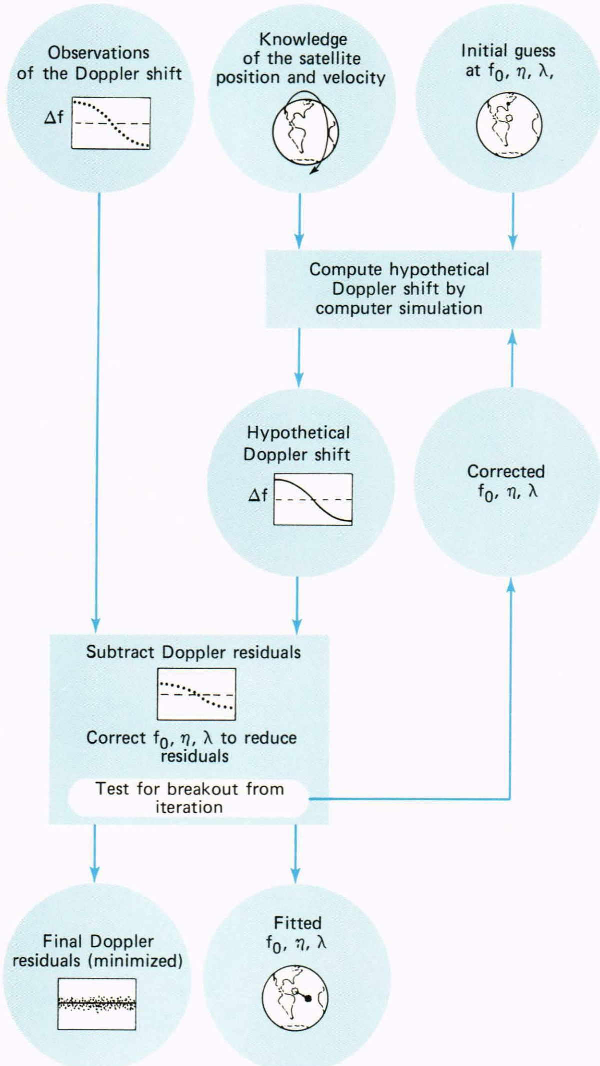

Fig. 6—Navigation computation algorithm. The technique is iterative, nonlinear least-squares. Three parameters are solved for: latitude (η), longitude (λ), and frequency bias (f_0). In the surveying mode, when multiple passes are used, distance from the earth's center is added and usually also a receiver bias. For the orbit determination problem, the same type of iterative algorithm is used.

MEASURING THE DOPPLER SHIFT — THE SATELLITE CASE

Thus far we have dealt with the Doppler shift as a measured frequency shift that is proportional to the range-rate, $d\rho/dt$ (see Eq. 1). In the train whistle analog, we were dealing with sound waves whereas for satellites we are concerned with electromagnetic radiation (a radio wave) that travels at nearly the speed of light. Consequently, we must reinterpret c. The underlying physical basis of the satellite system remains:

$$f = f_T + \Delta f = f_T\left(1 - \frac{1}{c}\frac{d\rho}{dt}\right) . \qquad (8)$$

Ideally, to isolate the Doppler shift for measurement we would subtract the received frequency, f, from a site-located oscillator tuned precisely to f_T. Mixing followed by low-pass filtering is a common technique for subtracting. The process would isolate $(f_T/c)\cdot(d\rho/dt)$. It is impossible to do exactly that. We must simultaneously cope with several other problems. First, neither the satellite nor the site oscillator is ideal, and both slowly change frequency for roughly the same reason that a wristwatch slowly loses or gains time. Second, the satellite signal, in reaching the observer, travels through the ionosphere and troposphere (the near atmosphere). Both interact with the frequency in ways that we describe in the next section. Third, if we electronically isolated the Doppler frequency, it would pass through zero frequency at the point of closest satellite approach. Very low frequencies are difficult to process electronically. However, the problem is easily circumvented by deliberately biasing the site oscillator. The sampled quantity becomes

$$f_M = -\frac{f_T}{c}\frac{d\rho}{dt} + \sum_{i=1}^{4} \delta f_i , \qquad (9)$$

where:

- δf_1 is a frequency bias term — a deliberately introduced term plus an unavoidable "drift" that changes slowly with time. If the site/satellite oscillators are carefully designed, we can treat δf_1 as a constant over the 15 to 20 minutes that the satellite spends above an observer's horizon. As a consequence, δf_1 can be removed in the data processing;
- δf_2 is the contribution of the nondispersive, neutral troposphere. About 90 to 95% of this effect can be removed by utilizing existing physical models of the tropospheric index of refraction (see below);
- δf_3 is a noise term that arises from a number of sources — nonideal oscillators and instrumentation. $\delta f_4/f_T$ is currently about 3×10^{-11}; and
- δf_4 is a contribution to the frequency (instantaneous phase) of the received signal caused by the ionosphere. δf_4 can be removed, sneakily, by taking advantage of the dispersive property of the ionosphere (see below).

The measurement of f_M proceeds by counting a fixed number (N) of cycles and recording the time required to obtain this count. Much care is exercised in this process to minimize the measurement errors. Typically a 5 MHz digital clock is (intermittently) read when N equals a preassigned value. (The ionospheric refraction error is removed before the counting is performed.) The measurement then is an integral of Eq. 9 (Fig. 7):

$$N(t_i, t_{i+1}) = \frac{f_T}{c}\left[\rho(t_{i+1}) - \rho(t_i)\right] + \delta f\left[(t_{i+1} - t_i)\right] . \qquad (10)$$

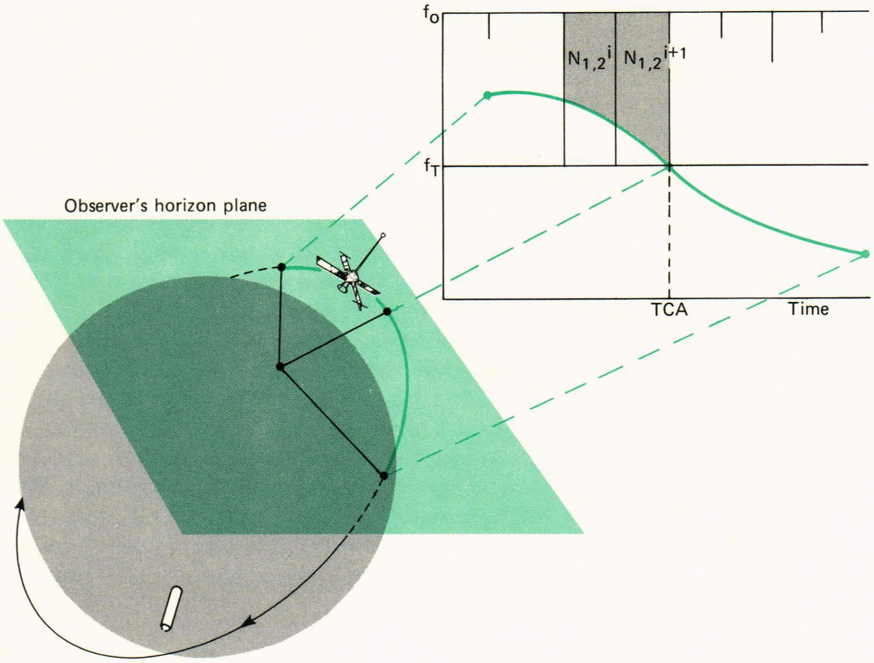

Fig. 7—Doppler-cycle count. The count is periodically repeated to sample repetitively (the integral of) the difference between a local oscillator (f_0) and a Doppler-shifted satellite signal.

(δf is the sum of the first three terms in the summation of Eq. 9.)

This same set of measurements is made both by people using the system (navigators and surveyors) and by the four tracking sites where data are collected for orbit determination. The important term in Eq. 10 is the slant range difference $\rho(t_{i+1}) - \rho(t_i)$. This is the geometric quantity that contains the satellite orbit on the one hand and the navigator's or surveyor's position on the other.

REMOVING PROPAGATION EFFECTS

If, as is certainly the case, the space between the satellite and the observer has something other than a homogeneous refractive index, then Eq. 8 must be generalized. We have developed the appropriate generalization on pages 10 and 11.

The Ionosphere

For the ionosphere, the necessary understanding is summarized in Eq. R7 on p. 11:

$$\Delta f = - \left\{ \frac{f_T}{c} \frac{d\rho}{dt} + \left[\delta f_4 = \frac{a_I}{f_T} + \frac{a_3}{f_T^{\,3}} + \ldots \right] \right\}. \tag{11}$$

This equation is the basis for eliminating the effects of ionospheric refraction. If we imagine arbitrarily raising the frequency broadcast by the satellite, then the ionospheric effect, δf_4, rapidly diminishes compared to the vacuum Doppler shift, $-f_T/c\ d\rho/dt$. It proved impractical to completely suppress the ionospheric terms so a compromise was struck: A pair of coherently related frequencies[4] is broadcast by the satellite and the pair is high enough (150 and 400 MHz) to suppress all but the first term of δf_4, the so-called "first order" ionospheric correction a_1/f_T. We can then construct a pair of simultaneous equations,

$$f^{(a)} = - \left[\frac{f_T^{\,(a)}}{c} \frac{d\rho}{dt} + \frac{a_I}{f_T^{\,(a)}} \right], \tag{12}$$

$$f^{(b)} = - \left[\frac{f_T^{\,(b)}}{c} \frac{d\rho}{dt} + \frac{a_I}{f_T^{\,(b)}} \right], \tag{13}$$

at every measurement time. We can solve this pair to create an expression that is linear in $d\rho/dt$ and that eliminates the a_1 term. The necessary algebra is performed in the data-gathering instrumentation. In designing the instrumentation, it is important to take advantage of the fact that $f_T^{\,(a)}/f_T^{\,(b)}$ is precisely $400/150 = 8/3$. Even though the frequencies change slowly with time, they change coherently, i.e., they maintain this fixed ratio.

This technique is not perfect as it leaves the higher order terms, principally $a_3/f_T^{\,3}$, as uncorrected biases in the data. The 400/150 MHz choice of frequencies is high enough to assure that the biases are rarely as large as 5 meters and typically less than 1 meter.

The Troposphere

Unlike the ionosphere, the tropospheric refractive index is not frequency dependent but, rather, depends — at every point — on the pressure/temperature ratio at that point (see p. 10). When the satellite is at an observer's zenith, the apparent (electromagnetically measured) range to the satellite is increased about 2⅓ meters. At larger zenith angles, the effect increases, but we can compute it quite accurately simply by knowing the satellite-station geometry and the surface pressure. There is an exception: the water vapor in the atmosphere has a small effect, typically 20 cm at the zenith, that can be only roughly compensated for because the distribution of water vapor in the atmosphere cannot be accurately modeled. Water vapor corrections present a fundamental limitation to space Doppler shift or range measurements.[1] We have outlined the analysis of tropospheric refraction on p. 10.

COMPUTING THE SATELLITE ORBIT

The satellite position is described by (and must obey) Newton's laws of motion, which specify — if the earth were spherical — that

$$\ddot{\bar{r}} = - \frac{GM}{r^2} \hat{r} \ , \tag{14}$$

where $\bar{r}$ is the satellite position, $r = |\bar{r}|$, $\hat{r} = \bar{r}/r$, $\ddot{\bar{r}} \equiv d^2(\bar{r})/dt^2$, and GM is the gravitational constant times the mass of the earth.

The position and velocity of the satellite at some one instant and the correct differential equation are, in principle, sufficient to specify the position of the satellite for all times. Neither of these — neither a correct motion equation nor a set of explicitly correct initial conditions — exists in reality.

Equation 14 specifies the familiar Kepler ellipse; a slightly more complicated representation of the motion is accurate to about 1 km for about one day. To represent the motion more precisely, we must replace the elementary (spherical) gravitational model of the earth with a more accurate model and add (a) the drag force caused by the motion of the satellite through the tenuous atmosphere, (b) a force caused by the sunlight, and (c) the gravitational forces caused by the sun and moon. To represent the gravity forces of the real earth requires some 450 terms. The basic principles and techniques associated with orbit determination do not change with this added complexity although the accuracy of the result improves from 1 km to about 1 meter.

The orbit determination technique proceeds as follows. The ionosphere-corrected data are $j = 1,2,3,\ldots J$ sets (passes) of N-values (Eq. 10) and times (t_i) recorded at the four tracking sites. The orbit is derived by minimizing the function

$$F(\bar{r}_0, \dot{\bar{r}}_0, \Delta f_j) = \sum_i \left[N(t_i) - N(\bar{r}_0, \dot{\bar{r}}_0, \Delta f_j, t_i) \right]^2 \tag{15}$$

with respect to the orbit initial conditions ($\bar{r}_0$, $\dot{\bar{r}}_0$) and the frequency bias parameters Δf_j.

This is the classical least-squares fit criterion. The numerical details are horrendous and involve numerically integrating the differential equation that describes the satellite position, a host of necessary accessory computations to compute N from Eq. 10, and a procedure for minimizing F. Guier and Weiffenbach in their original paper[3] showed conclusively that this computation was practical and did produce an accurate description of the satellite orbit. It was crucial to the success of the technique that the frequency bias parameters, Δf (one for each pass of the satellite over each tracking site), be included among the solved-for orbit parameters. The logic associated with the computation is similar to that in the navigation computation (Fig. 6) albeit more complicated because of the larger number of solved-for parameters.

The navigator has a much easier computation. She or he has the satellite position available and solves for (minimizes F with respect to) one frequency bias parameter plus latitude and longitude (Fig. 6).

SYSTEM ACCURACY AND PRECISION

Precision is the reproducibility or internal consistency of a measurement. To be accurate, the system must be consistently precise, but this is not enough. Accuracy is a more stringent criterion that requires the acceptance of an absolute standard and comparison with that standard.

There are two useful measures of system precision: one for navigational use and another for surveying. To obtain either one, we take repeated data samples (passes of a satellite) at a fixed site. The navigational precision is obtained by using the ephemeris from the satellite to navigate (compute latitude, longitude, and frequency bias) for each individual pass. The individual latitude-longitude points are plotted. Figure 8 is a typical example. It shows that the navigator's usual error, since he uses a single pass of data for his results, is about 25 meters. This, however, is deceptively small for the at-sea navigator. When used under way, the system is less precise than it is on land at a fixed point. The uncertainty of the ship's speed (not the speed itself) introduces a bias into the data that cannot be removed. For every 1 knot error in the ship's speed, the associated position error is about 400 meters. This is not important for most ships. It is not difficult to know the ship's velocity with an accuracy of 1/4 to 1/2 knot. The associated 100 to 200 meters position error is comparable to the length of the ship and is of no concern to most ship captains.

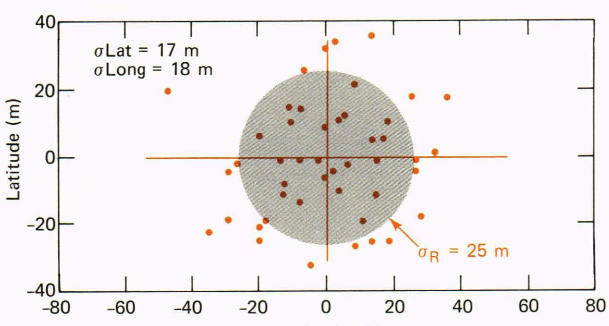

Fig. 8—Navigation results obtained at a fixed site in real time. The individual data points are independent navigation computations. A navigator at sea would get a single point about every 1½ hours.

The surveyor, in contrast, is not so much interested in the individual pass results as he is in the accuracy of the combined results. Current experience[5,6] gives ½ to 1½ meters for this accuracy. The surveyor can get this result in either of two ways. He can, in real time, use the ephemeris from the satellite and obtain individual results typical of those in Fig. 8. The center of the bullseye of 50 passes would have an uncertainty of about $20/\sqrt{50} \simeq 3$ meters (the mean of another 50 passes would agree with the previous one within about 3 meters). Alternatively, rather than using the real-time-predicted ephemeris from the satellite, he can use an *ex post facto* computed ephemeris. The latter approach has the apparent advantage that it obtains an equivalent result and requires less data. The predominant error in the satellite-borne (extrapolated) ephemeris is caused by the atmospheric density at satellite altitude, which is difficult to predict. We can rid the ephemeris of the error source if we compute *ex post facto* the ephemeris using data that span the time interval during which the survey is being performed. An example is shown in Fig. 9. Elementary statistics would say that the precision of the mean surveyed position is 0.4 meters; however, repeated attempts show that it is typically a few tens of centimeters to 1.5 meters.[5,7,8] This is the (so-called) absolute or global accuracy of the system as used for surveying.

The more commonly used method is to determine simultaneously the relative positions of several points within one region. This is the "intervisible" mode. If a satellite is visible at several sites simultaneously, errors in satellite position affect the surveyed position of the sites in a nearly identical way. As a conse-

CORRECTING FOR TROPOSPHERIC AND IONOSPHERIC (REFRACTIVE) EFFECTS

Since the path from the satellite to the observer traverses both the ionosphere and the troposphere, both of which affect the phase of the transmitted signal, we can include these effects in the theory by replacing the "vacuum-path" formulation

$$\Delta f = -\frac{f_T}{c}\frac{d\rho}{dt},$$

with

$$\Delta f = -f_T \frac{d}{dt}\int_p \frac{d\rho}{u(\rho,t)}, \quad \text{(R1)}$$

where u is the phase velocity of the signal. This varies at every point within the ionosphere. If we introduce the index of refraction, n,

$$n \triangleq \frac{c}{u(\rho,t)}, \quad \text{(R2)}$$

in Eq. R1,

$$\Delta f = -\frac{f_T}{c}\frac{d}{dt}\left[\Delta s = \int_p n(\rho,t)\,d\rho\right] \quad \text{(R3)}$$

where c is the speed of light.

There is an impossible-to-know requirement here. The line integral (the optical path length) must be evaluated along the extremum path (p) in accord with Fermat's principle. Since we do not know n at every point, this is impossible. It is also unnecessary; because n is so close to unity, the instantaneous straight line connecting the satellite and observer suffices. The tropospheric effects peter out about 40 km above the surface, whereas the ionosphere begins above 80 km and extends upward. Moreover, these two regions have distinctly different effects on the phase of the broadcast signal. The effects can be characterized by the way each region affects the refractive index. For the troposphere, $n > 1$, whereas for the ionosphere $n < 1$. Since either slightly alters the phase of the signal, we can treat the effects independently and superimpose their effects.

THE TROPOSPHERE

The index of refraction at any point within the troposphere is given by the Smith/Weintraub expression

$$n = \frac{77.6P}{T} + 3.73\times 10^5\,\frac{e}{T^2}$$

wherein P and T are the pressure in millibars and temperature in K at a point, and e is the partial pressure of the atmospheric water vapor, in millibars. The first term, because it is independent of the water vapor, is called the "dry" term. It is typically ten times as large as the "wet" term.

If, as is typically true, the temperature decreases linearly with height, then the optical path length through the (dry) troposphere is the vacuum path plus Δs (see Refs. R1, R2, and R3):

$$\Delta s = 2.343\,P_s\left[\frac{T-4.12}{T}\right]\cdot I_a,$$

$$I_a = \left\{1 - \left[\frac{\cos^2 E}{1+0.15\,h_d/r_s}\right]^2\right\}^{-1/2}. \quad \text{(R4)}$$

wherein Δs is the dry-tropospheric range correction in meters; P_s is the surface pressure in standard atmospheres; T is the surface temperature in K; E is the elevation angle — the angle between the observer-satellite line and the horizon; r_s is the distance from the earth center to the observer; h_d, the dry-tropospheric "extent," equals 148.98 (T − 4.12). h_d is in the 34 to 40 km range.

Computations show that the temperature dependence is unimportant for $E \geq 5°$ and, moreover, for $E \geq 30°$, Eq. R4 reduces to

$$\Delta s = 2.31\,P_s \operatorname{cosec} E. \quad \text{(R5)}$$

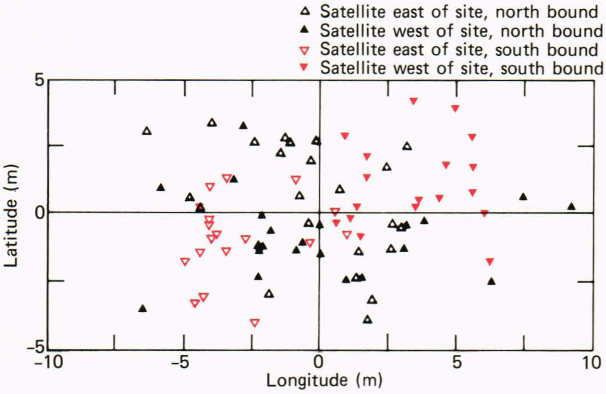

Fig. 9—Results of a recent surveying computation. The elongated data point distribution is exploited to determine the station radius. A 1 meter bias was deliberately introduced here to illustrate the effect.

quence, the error does not affect their relative positions, which are then known more precisely (10 to 40 cm) than is any point absolutely (globally).[7]

Using multiple passes, it is quite practical to determine all three coordinates (latitude, longitude, radial distance from the earth's center) of a fixed site. The data shown in Fig. 9 illustrate the effect of an incorrect radial distance. Had the distance been precisely correct, the data point distribution would have been more nearly circular rather than so elongated in longitude. A closer examination of the data shows that there are two distributions of data, one characterized by the satellite passing east of the observing site and another by the satellite passing west. When the two distributions completely overlap, the radial distance is correct. The data in Fig. 9 correspond to the site radius being about 1 meter too long. In the usual

Equation R5 contains the dominant functional effects of the troposphere on the satellite signal. For a satellite directly overhead and for a surface pressure of 1 atmosphere (it hardly varies from 1 atmosphere unless the observer is in a hurricane or on a mountain), the tropospheric effect is 2.31 meters. The wet term adds about another 20 cm.

The necessary understanding that culminated in this model was developed by Helen Hopfield of APL over a 15-year period (Ref. R1).

The wet term is in a much less satisfactory state; some recent progress and the current state are described in a recent paper by Goldfinger (Ref. R2).

THE IONOSPHERE[R4]

The index of refraction of the ionosphere is, for our purposes, given by a simplified form of the Appleton-Hartree formula:

$$n = \left[1 - \left(\frac{f_N}{f_T}\right)^2\right]^{1/2} . \qquad \text{(R6)}$$

The validity of this simplified form requires that the satellite frequency be high compared with the electron plasma resonance frequency, f_N. A typical value of f_N is 10 MHz. f_N varies from point to point within the ionosphere and is (in hertz)

$$f_N = \sqrt{80.6N}$$

where N is the local electron density in electrons/m^3.

Since $f_N \ll f$, we can expand Eq. R6 in a rapidly converging series

$$n = 1 - \frac{1}{2}\left(\frac{f_N}{f_T}\right)^2 - \frac{1}{8}\left(\frac{f_N}{f_T}\right)^4 + \ldots ,$$

and substitute into Eq. R2,

$$\Delta f = -\left[\frac{f_T}{c}\frac{d\rho}{dt} + \frac{a_1}{f_T} + \frac{a_3}{f_T^{\,3}} + \ldots\right] . \qquad \text{(R7)}$$

We have not written out the definitions of a_1, a_3, etc. because it is only necessary to know that they are independent of the satellite frequency. We truncate this series after the first-order (a_1) term. The 150/400 MHz frequency pair is high enough so that the a_3 term is less than 1/10th that of the [first order] a_1 term. The first-order ionospheric correction (the correction to the data shown in Fig. 1) is shown in Fig. R1.

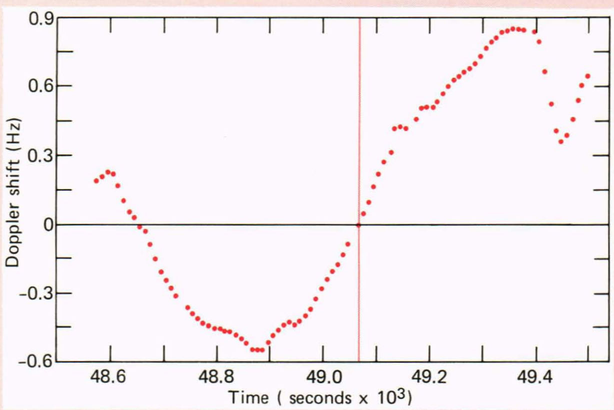

Fig. R1—The first-order ionospheric correction to the data shown in Fig. 1. The correction is of the order of 1:10^4 of the Doppler shift.

REFERENCES

R1 H. S. Hopfield, "Tropospheric Effect on Electromagnetically-Measured Range: Prediction from Surface Weather Data," *Radio Sci.* **6**, pp. 357-367 (1971).

R2 A. D. Goldfinger, "Refraction of Microwave Signals by Water Vapor," *J. Geophys. Res.* **85**, pp. 4904-4912 (1980).

R3 H. D. Black, "An Easily Implemented Algorithm for the Tropospheric Range Correction," *J. Geophys. Res.* **83**, pp. 1825-1828 (1978).

R4 W. H. Guier, and G. C. Weiffenbach, "A Satellite Doppler Navigation System, *Proc. IRE* **48-4**, pp. 507-516 (1960).

survey computation, this distribution bias is used to determine simultaneously the radius along with latitude and longitude. A more basic description of this process is contained in Ref. 9.

Determining the accuracy of the system is a very subtle and difficult problem because there has been no accepted accuracy standard for global surveying or navigation. Within the borders of individual countries, there have been recognized standards (e.g., the North American Datum), but globally none existed. "Ties" (coordinate transformations) between the various datums were not uniformly reliable, particularly for regions remote from North America or Europe. For example, we repositioned Australia and Hawaii several hundred meters relative to North America in the mid-1960's, when the system was reliably producing surveying accuracies. Similarly, a number of remote islands (e.g., Ascension) have had to be repositioned by similar amounts for the following reason. Positions of remote islands heretofore have relied strongly on optical measurements of angles to stars (Fig. 10). The interpretation of these angular measurements depends on knowing the reference directions at the two points (local or "plumb-bob" verticals) in a common coordinate system. The local plumb-bob direction is changed by subsurface density variations within the earth. As a consequence, there were inconsistencies in the interpretation of the astronomical readings. Geodesists call this the "deflection-of-the-vertical" problem. The Transit system provided a technique that, for the first time, was immune to these vagaries; Transit does not depend on angular or directional measurement.

There was no surveying standard that had higher precision than did Transit, certainly not for long over-the-water surveying. As a consequence, the ac-

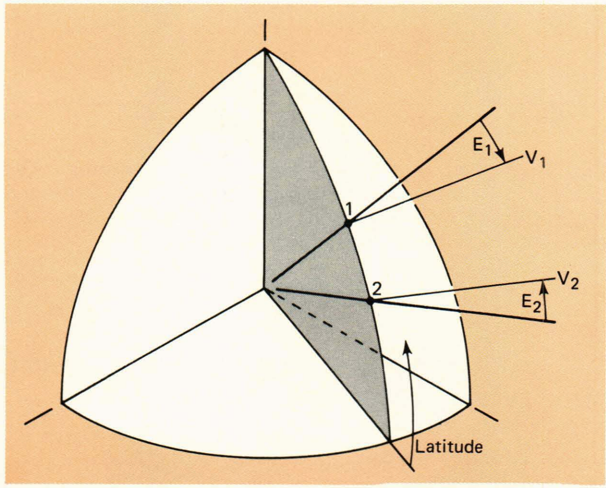

Fig. 10—Determining latitude from astronomic measurements. Errors caused by uncertainties in the direction of the local plumb-bob verticals (the lines V_1 and V_2) cannot be removed using classic techniques. The latitude errors at the two sites are E_1 and E_2.

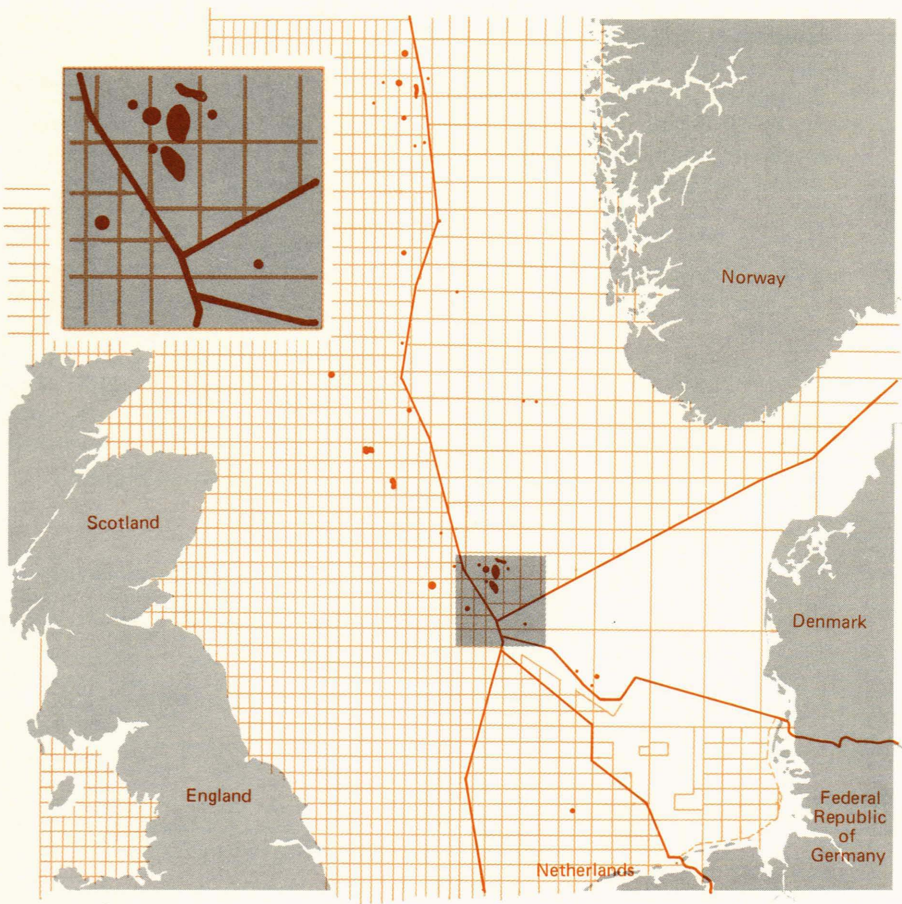

Fig. 11—Oil wells and oil fields in the North Sea. Nearly all were located using the Transit satellites. Grids indicate potential lease sites, and heavy boundary lines represent limits of national drilling rights. Oil wells and oil fields are shown as points or random shapes, respectively.

curacy question was moot; there was no accepted standard for comparison. In another sense, the precision available from the system provides a limited measure of accuracy. This is true because the system depends on several fundamental constants, e.g., the mean sidereal rate of the earth ($\omega_e = 7.292115855 \times 10^{-5}$ rad/s) and the speed of light ($c = 299{,}792.5 \pm 0.05$ km/s)[10] which are supplied from external sources; these constants are, in turn, tied directly to fundamental standards of length and time. Experiments show that the accuracy of position determination is 1 to 5×10^{-7} of the radius of the earth.[5,6] A related fact is that the accuracy of c is about 1 part in 10^{7} (Ref. 11). The *GM* we are currently using is now known to be 1 part in 10^6 too large.[12]

There are nonrandom errors in Transit (e.g., drag and geopotential model errors) in addition to the errors in the basic constants. Consequently, we cannot make a definitive analysis or measure of the system's accuracy on a global basis. Over distances of several hundred to a thousand kilometers, when compared with direct distance measures, Transit is accurate to a few decimeters (see above).

USERS AND USER EQUIPMENT

There are currently about 16,000 users of the Transit system.[13,14] There are 17 manufacturers of navigation and surveying equipment, both in the U.S.A. and abroad. Most of the navigation sets are single-frequency receivers that do not correct for ionospheric effects and that accept the 100 to 200 meters associated error. The cost of the single frequency sets is less than half that of the more accurate receiver. The least expensive sets sell for $4,000. The more accurate, dual-frequency sets are typically $20,000 to $30,000.

Transit has met a crucial need in positioning offshore drilling platforms to assure that rigs were located within the leases and that they did not violate international boundaries. This problem was particularly important in the North Sea where drilling rights are allocated to bordering countries (Fig. 11).

REFERENCES and NOTES

[1]W. H. Guier and G. C. Weiffenbach, "The Early Days of Sputnik," *Johns Hopkins APL Tech. Dig.* (this issue).

[2]R. R. Newton, "The Near-Term Potential of Doppler Location," *Johns Hopkins APL Tech. Digest* (this issue).

[3]W. H. Guier and G. C. Weiffenbach, "Theoretical Analysis of Doppler Radio Signals from Earth Satellites," JHU/APL BB-276 (1958).

[4]The pair is derived from a common oscillator and then carefully processed to preserve their fixed-phase relationship.

[5]C. Boucher *et al.*, "The Second European Doppler Observation Campaign (EDOC-2)," *Proc. 2nd International Geodetic Symposium on Satellite Doppler Positioning,* Austin, TX, pp. 819-849 (1979).

[6]J. Kouba, "Improvements in Canadian Geodetic Doppler Programs," *Proc. 2nd International Geodetic Symposium on Satellite Doppler Positioning*, Austin, TX, pp. 63-82, (1979).

[7]P. Wilson, "On Some Geodetic Uses of the Navy Navigation Satellite System by European Agencies," *Proc. National Aerospace Symposium*, The Institute of Navigation, Washington, D.C., pp. 101-110 (1979).

[8]V. Ashkenazi *et al.*, "Terrestrial-Doppler Adjustment and Analysis of the Primary Triangulation of Great Britain: Preliminary Report," *Philos. Trans. R. Soc. London* **A 294**, pp. 385-394 (1980).

[9]H. D. Black, "The Transit System, 1977: Performance, Plans and Potential," *Philos. Trans. R. Soc. London,* **A 294**, pp. 217-236 (1980).

[10]It is an interesting fact that the way the international meter is defined currently places (or did place in 1975) a limitation on the ability to measure precisely the speed of light.[11] Evenson gives $c = 299{,}792.458 \pm 0.001$.

[11]K. M. Evenson, "The Development of Direct Optical Frequency Measurement and the Speed of Light," *ISA Trans.* **14**, pp. 209-216 (1975).

[12]D. E. Smith, "A Determination of GM," NASA Tech. Memo.80642, pp. 23-26 (1980).

[13]T. A. Stansell, "The Transit Navigation Satellite System," Magnavox Report R-5933 (1978).

[14]T. A. Stansell, "The Many Faces of Transit," *NAVIGATION: J. Inst. Nav.* **25**, No. 1, pp. 55-70 (1978).

ACKNOWLEDGMENTS—I am indebted to R. B. Kershner, W. H. Guier, W. L. Ebert, and G. C. Weiffenbach for helpful discussions; to M. M. Feen for the ionospheric data; and to the Navy Astronautics Group of Point Mugu, California, who operate the Transit System.