History of Key Technologies

Early Development of Transit, The Navy Navigation Satellite System

Harold D. Black
Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland 20723

I. Introduction

WE owe the Soviet people our sincere thanks for the first artificial satellite, Sputnik I. Their intellectual gauntlet, "thrown up" in October 1957, challenged our political and technical leadership. Responses to this challenge have enriched the lives of us all through improved global communication, weather forecasting, ship navigation, and land surveying. As a result of Projects Apollo and Voyager, we have an expanded sense of our own horizons.

In the immediate aftermath of Sputnik, the Transit system (a.k.a. Navy Navigation Satellite System, NNSS, NAVSAT) was invented by F. T. McClure, W. H. Guier, and G. C. Weiffenbach. Under the leadership of R. B. Kershner, the system was developed at the Johns Hopkins University Applied Physics Laboratory over the six-year period of 1958–1964.

When the space division of the laboratory was formally established in January 1960, I was thrust into a responsible role for analysis and programming: in short, to make the software work.

Transit was placed in operation in mid-1964. A presidential decree released the system for public use in 1967 and, since that time, about 80,000 navigation sets have been sold by a number of private companies. The system is in intensive daily use and has been for the last 25 years.

The system pioneered the use of satellites for ocean navigation and land surveying and reduced the arts of surveying and navigation to a solved technical problem with a firm scientific basis. Recognizing this, in 1986 the Institute for Electrical and Electronics Engineers (IEEE) awarded the inventors (McClure, Guier, and Weiffenbach) and R. B. Kershner The Pioneer Award (see *IEEE Transactions on Aerospace and Electronic Systems,* Vol. AES-22, July 1986).

This paper is intended to sketch how the system works—its undergirding principles—and to describe the early (1958–64) development of the system.

The literature on Transit is vast. An annotated bibliography is included in Ref. 1. Reference 2 is a reasonably complete bibliography for the period 1958–1979. Also, one entire issue of the *Johns Hopkins APL Technical Digest* (Vol. 2, No. 1, 1981) is a collection of papers on the Transit system.

II. The Problem

Classical (celestial) navigation and surveying have both fundamental and practical limitations that limit their accuracy and utility: A celestial navigator cannot measure the altitude (elevation angle) of a star when the sky is overcast or in the daytime; nor can he measure precisely while his ship is rolling and pitching. The traditional surveyor is similarly limited but for more subtle reasons. The surveyor also measures angles; his reference direction, "the plumb bob vertical," is corrupted by local anomalies in the gravitational field. Moreover, the surveyor is unable to determine the absolute distance from the Earth's center to a point on the surface. "Leveling" provides relative height information. Practically, surveyors are hampered by oceans, national boundaries, and difficult terrain.

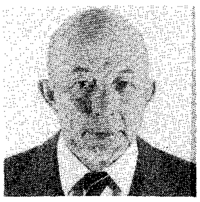
The Transit system removed these restrictions: it provided 1–5-m accuracy in absolute global position and established the basis for continuously improving navigation and surveying. This basis is characterized by 1) accuracy improvements in satellite position provided by better system "clocks" (oscillators) and improved models for the satellite motion, 2) associated improvements in ground station equipment, and 3) improved methods of dealing with ionospheric and tropospheric effects on the satellite signal. We will have more to say on this subject.

III. Invention of the System

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: "After listening to all the passes that first week it was apparent that the Doppler shift was also a distinctive and reliable signature for Sputnik...."[3]

They heard a characteristic drop in pitch as the satellite approached and then receded into the distance. This "Doppler shift" is familiar to anyone who has waited at a railroad crossing and listened to the whistle as the train approached and then disappeared down the track. A characteristic Doppler curve is shown in Fig. 1.



Harold D. Black was a member of the Principal Staff of the Johns Hopkins University Applied Physics Laboratory. He grew up in Concord, North Carolina, and was educated at North Carolina State University (BME) and at Carnegie Institute of Technology (MS). For a number of years he was the Group Supervisor of the Space Analysis and Computation group in the Space Department of JHU/APL. From 1986 to 1988, he was on loan to the U.S. Naval Academy where he occupied the Naval Space Command Chair in the Aerospace Engineering Department. He retired in November of 1989 and his hobbies are gardening and classical music.

EDITORS'S NOTE: This manuscript was invited as a History of Key Technologies paper. It is not meant to be a comprehensive study of the field. It represents solely the author's own recollection of events at the time and is based upon his own experiences.

Received Oct. 17, 1988; revision received Jan. 10, 1989.

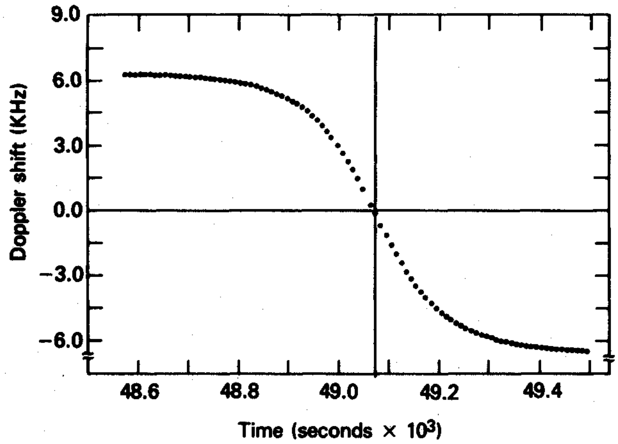


Fig. 1 Doppler shift of a 300 MHz satellite oscillator in an 1100 km orbit. The observing site is at 45° latitude, and the satellite is 77° above the observer's horizon at the time of closest approach. Zero Doppler shift occurs at that time.

After an intensive effort, Guier and Weiffenbach showed that the shape of this curve, when used together with the laws of motion, implicitly contained the satellite orbit. The Doppler shift was a familiar phenomenon. But when they tried to match (least squares fit) the shape of the measured curve with successively more accurate orbit theory, the data could not be accurately fitted until a rigorously correct description of the satellite motion was used. Then there was a match between the key and lock: only one orbit would fit the data.

It was not as simple as this sounds. Guier and Weiffenbach made a number of original discoveries and innovations. For example, they recognized that meaningful frequency measurements required that the satellite-borne oscillator and the ground oscillators must be calibrated against a common frequency standard (WWV). This was impossible using conventional techniques because both oscillators changed slowly with time and the satellite oscillator was in orbit. They discovered that they could perform the calibration with the satellite in orbit by utilizing the entire Doppler curve and by including the calibration, frequency bias terms, in the orbit state vector.[4]

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

The navigation system was invented when the late F. T. McClure, then Chairman of the APL Research Center, realized that the discovery of Guier and Weiffenbach could be "inverted." If they could derive the satellite orbit by using Doppler data obtained at *known* sites, then, using that orbit, navigators at *unknown* sites could, with Doppler shift measurements, derive their position coordinates. The APL staff recognized that the system should have some striking advantages over all existing forms of navigation: it was all-weather, it would work both day and night, and it would not require directional antennas.

IV. System Description

Figure 2 shows schematically the system architecture and its four basic elements: 1) A satellite constellation, a single satellite would suffice; 2) a set of four tracking stations and a computation center for processing the data; 3) stations for injecting an ephemeris specification into the satellites; and 4) users.

Constellation of Satellites

A number of operational satellites (seven in 1990) are in near-Earth orbits that pass over the Earth's poles. The orbits are nominally circular at an altitude of approximately 1100 km. Each satellite contains the following:

1) A highly precise frequency standard (oscillator) that "drives" two transmitters, nominally at 150 and 400 MHz. (The 1960 electronic technology strongly influenced the 400 MHz choice.) Both frequencies are "coherent"; because they are derived from a common oscillator, a physical constraint exists between their phases. A cycle counter driven by this same oscillator serves as a digital clock.

2) A memory that holds a current "satellite ephemeris specification," the necessary information to construct satellite position and velocity vs time. The ephemeris information and the clock control register can be revised from the ground by means of a communication channel. One axis of the satellite is oriented along the local vertical; the antennas always face the Earth. The ephemeris information and satellite identification number are continuously relayed (at 50 baud) to the navigator via modulation patterns superimposed on the 150 and 400 MHz transmissions.

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 received satellite signal at 4-s intervals. After the satellite has set (typically, 15 min 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 determine a contemporary orbit specification for each of the satellites and prepare an ephemeris for the next 24 h. As a part of this same computation, all system oscillators (clocks) are calibrated relative to a common standard. The measurements are also used to compute the necessary satellite clock corrections to compensate for the predictable part of the oscillator drift, typically $\Delta f/f = 10^{-11}$ parts per day. The predicted ephemeris specification and the satellite clock correction information are then relayed to one of three injection sites.

Injection Station

The new ephemeris and the clock correction are injected into the satellite memory via a communication channel, each injection writing over the one that is about to expire. Injections are at 12-h intervals; every satellite is visible at every station at least once every 12 h. The satellite memory has sufficient storage to contain a 16-h ephemeris.

User

A user's navigation set measures the received satellite frequency at discrete intervals and recovers the satellite ephemeris information broadcast by the satellite. With the frequency measurements, the satellite ephemeris, and the navigator's own motion, the position is automatically computed.

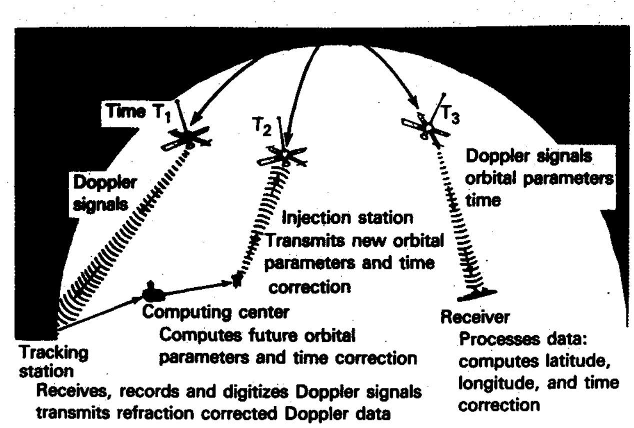


Fig. 2 System architecture of the Navy Navigation Satellite System (Transit).

Although 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 brevity's sake, we omit a number of details that can be found in the references.

V. Positioning

There are two kinds of position determination: navigation and surveying. These differ in several respects. The navigator is not concerned with errors that are smaller than the length of his moving ship, nor is he concerned with his distance from the Earth's center. Surveying establishes the position of an Earth-fixed point, usually in three dimensions and as accurately as possible. Surveyors establish an antenna site and collect many passes of data for subsequent analysis. The navigator needs a position immediately.

VI. How the System Works

Descriptions of how the system works—why the orbit is determinate from the Doppler shift measurements and why, given the satellite ephemeris, the navigator's position is determinate—exist on several different levels. The dependence of the Doppler shift on the satellite-observer position and velocity is important in understanding them all.

For any but the most precise uses, we can write the Doppler shift (see the Appendix for derivation and notation) as

$$\Delta f = -\frac{f_s}{c}\frac{\mathrm{d}\rho}{\mathrm{d}t} \tag{1}$$

wherein ρ, the "slant range," is defined as

$$\rho \triangleq |\bar{r}(t^*) - \bar{r}_T(t)| \tag{2}$$

and

$$t^* \triangleq t - \rho/c \tag{3}$$

As explained in the Appendix, Eq. (3) is the basis of associating epochs on two clocks, one in the satellite (t^*) and another (t) at the observer's site.

In measuring frequency, cycle counting (integration) is unavoidable. Rather than dividing the cycle count by the elapsed time interval to obtain frequency, it is convenient to use the integrated cycle count (as a function of time) as the observed data. Typically, the cycle count is repeated at 2-s or 4-s intervals. To clarify, if we subtract (mix and then low-pass filter) the incoming satellite signal from an observer-generated frequency f_o, and then integrate, we obtain

$$\int_{t_k}^{t_{k+1}}\left[f_o - (f_s + \Delta f)\right]\mathrm{d}t \equiv \int_{t_k}^{t_{k+1}}\left[f_o - \left(f_s - \frac{f_s}{c}\frac{\mathrm{d}\rho}{\mathrm{d}t}\right)\right]\mathrm{d}t \tag{4a}$$

or

$$_kN_{k+1} = \int_{t_k}^{t_{k+1}}\left(\frac{fs}{c}\frac{\mathrm{d}\rho}{\mathrm{d}t}\right)\mathrm{d}t + \int_{t_k}^{t_{k+1}}(f_o - f_s)\,\mathrm{d}t \tag{4b}$$

Since 1) the counting interval $(t_{k+1} - t_k)$ is a few seconds, and 2) the oscillator noise $\sigma(\Delta f/f)$ is 10^{-10} to 10^{-11} and the oscillators are stable over periods of a day (also 10^{-10} to 10^{-11}), we can simplify the previous equation into the form

$$_kN_{k+1} = (fs/c)(\rho_{k+1} - \rho_k) + \delta f(t_{k+1} - t_k) \tag{5}$$

wherein

$$\delta f \triangleq f_o - f_s \quad \text{(the frequency bias)} \tag{6}$$

Equation (5) is the basic observational equation.

Since neither the satellite nor the observer's oscillator are absolutely stable, we assign an independent δf to each "pass" (transit) of data. A pass is typically 15 min long.

Equation (5) is also the basis of understanding how measurements of the Doppler shift ("cycle count") provide positional information: Temporarily ignoring the δf-term, successive measurements of N define successive hyperbolic surfaces having foci at the satellite positions. Paraphrasing Kershner,[5] who had a deep technical understanding of the system:

> The Navy Navigational Satellite System is similar to an ordinary hyperbolic radio navigation system but one in which the role of the fixed ground network of transmitting stations is replaced by a sequence of successive positions of a single orbiting satellite. Suppose that at a given time, t_1, the satellite is at a given point, P_1. At a later time, t_2, the satellite has moved to P_2. Then if it were possible to measure the difference in distance from the navigator's (unknown) position to the two points, the navigator would be constrained to lie on the surface of a hyperboloid. If the difference in time is about 2 minutes, the distance from P_1 to P_2 is about 1000 km which is a reasonable baseline for a hyperbolic system. Waiting another 2 minutes gives a third position and a second baseline for another family of hyperboloids. The intersection of each determined hyperboloid with the surface of the earth gives a line of position, and the two lines of position intersect in a position determination. The problem is actually a little more complicated since we must deal with a frequency bias term. However, the complication only requires using multiple baselines and a least squares estimation process.

This explanation is useful in understanding that the navigator's position is determinate *without* measuring angles or distances per se. The navigator's computation is a nonlinear least squares estimator that (iteratively) produces the latitude ϕ, longitude λ, and frequency bias δf from minimizing

$$F(\phi, \lambda, \delta f) = \sum_k [N_k - N_k(\phi, \lambda, \delta f)]^2 \tag{7}$$

The N on the far right is computed using the right-hand side of Eq. (5), the current orbit specification, and models of the navigator's motion: his motion produced by the rotation of the Earth plus his course and speed.

VII. Deriving the Orbit

The orbit specification is produced in a similar fashion. The least squares estimator is parameterized in initial conditions for the differential equations describing the orbit, plus frequency bias terms. Producing the orbit is a daunting computation. Data from the four permanent tracking sites are accumulated about once a day and included with a portion of the previous day's data. The set of passes (typically 20, each consisting of several hundred measurements) are then processed through the orbit determination program. In writing the orbit determination software, we were simply lucky on several counts. The first version, ca. 1959, written in assembly language for the Univac 1103AF computer, would barely keep up with the data. Nevertheless it provided a test bed for the program logic and physics. The data noise estimates, a by-product of the least squares orbit determination, provided incentive to improve ground and satellite hardware, particularly oscillators. The structured (correlated) biases were quickly recognized as deficiencies in the gravity field model. In 1959 we obtained an IBM 7090 computer, about 10 times faster than the Univac 1103AF. After a whirl-wind reprogramming effort, we broke our computer-time bottleneck.

On another count, we had computer programmers who early appreciated the importance of *first* designing the programs and *then* programming them. Almost from the beginning we used the (now-called) "structured programming." Our software was highly modularized (closed subroutines), otherwise we would never have succeeded with the 100-person-year software effort. The first, reasonably bug-free version of the Transit orbit determination program was delivered to the Navy Astronautics Group, the operating agency, on Friday, December 13, 1963. The computer listing was about 1 m high and it required 1.5 h to execute on the IBM 7090. This program was used in early 1964 to certify that the system would provide 0.1 n.mi. navigation. This was the first time anyone had developed a technique for placing all terrestrial points on a rigorously consistent global datum.

Because the program was modularized with a central pool of constants, it was easy to maintain. The program evolved over the next 16 years. We changed it about every two years to keep pace with changes in the station hardware, to improve the algorithms, and to revise the "constants." The gravitational constant, the speed of light, the mean radius of the Earth, and the geopotential coefficient set all changed.

One of the many frustrations of that era was trying to understand the coordinate systems and conventions used in *The American Ephemeris and Nautical Almanac* (the *AENA*), now called *The Astronomical Almanac*. None of us had had any formal training in astronomy or celestial mechanics, but we did understand Eulerian rotation matrices, the convenient way to describe coordinate transformations. With its host of coordinate systems, there is no mention of a rotation matrix in the *AENA*. We waited with great anticipation for "The Explanatory Supplement to the *AENA*," which appeared in 1961. It was of little help on coordinate rotations.

In the pre-1960 days, the now familiar "A^TA" formulation of the least squares algorithm was unknown to us (Gauss knew it!). Using rare physical insight, W. H. Guier formulated several, increasingly more sophisticated, gradient-based minimization techniques for finding the orbit initial conditions (and frequency bias terms) from a least-squares estimation algorithm. He recognized and dealt successfully with "ill-conditioning" and the slow convergence of gradient searching.

VIII. Satellite Design

The original block diagram of the satellite is shown in Fig. 3.[6] Subsequent to this original plan, it was realized that many satellite parts would have to be redundant. The initial battery weight turned out to be correct, but the solar cell area increased to 5200 in.^{2} to provide for in-orbit degradation. (The operational, "Oscar," design weighs 110 lbs, 50 kg.) Attitude control and monitoring systems were also added. We will discuss the early development of each of these systems.

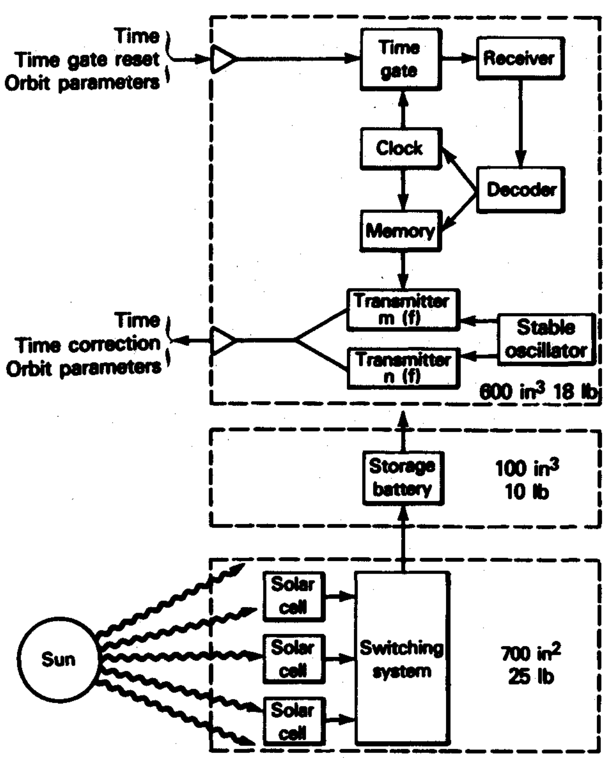


Fig. 3 Block diagram of the Transit satellite, 1959.

IX. Attitude Control System

Early in the program, the necessity for controlling the satellite attitude was recognized. *Any* unknown velocity of the satellite antenna (e.g., "tumbling") would increase the uncertainty of the observed Doppler shift. Moreover, without attitude control, the orientation relative to the sun would be unpredictable and the solar-cell-generated power for the satellite would be uncertain. While precise pointing was not required, stability and simplicity were. As Kershner[7] states:

> As a stopgap, our early satellites after Transit I-A were magnetically stabilized by incorporating a fixed magnet in the satellite that would cause it to line up along the earth's magnetic field lines. This had the virtue of pointing one specific side of the satellite toward the earth in the northern hemisphere. It is true that the other end would point earthward in the southern hemisphere, but at that time, no ground stations were located in the southern hemisphere. Magnetic stabilization served our purposes for a while.
>
> Some means of damping was required to make magnetic "stabilization" actually stable. At the suggestion of R. E. Fischell, this problem was solved very simply and ingeniously by the use of "hysteresis rods," i.e., long, thin rods of a readily magnetizable material that exhibit substantial magnetic hysteresis.... For every complete revolution of the rod, a fixed amount of energy is dissipated in the rod. This amount of energy is the area of the standard hysteresis loop. Thus, the rate of energy loss is proportional to the angular velocity, or
>
> $$\frac{d}{dt}\left(\frac{1}{2}I\omega^2\right) = -|\text{const}|\omega$$
>
> whence,
>
> $$I\dot{\omega} = -|\text{const}|$$

The angular velocity monotonically decreases. This technique is more effective than viscous damping, which diminishes in effectiveness as the velocity decreases. Moreover and most important, the technique is "passive"—no moving parts. The simplicity criterion is satisfied.

There are four hysteresis rods in the satellite. Since the area inside the hysterisis loop increases with the rod L/D ratio, each rod is a few tens-of-centimeters long and 0.3 cm in diameter. The rods are fabricated from a high permeability Ni/Fe alloy.[8] The characteristic time to reach zero angular speed is inversely proportional to the spin axis moment of inertia.

Although magnetic stabilization was replaced with "gravitational gradient" stabilization, the elegance of magnetic damping has persisted and has been widely employed in many other spacecraft.

X. Gravitational Gradient Stabilization

Early in the program we recognized[9] that the gravitational torque produced on an asymmetric body was a means for passive attitude stabilization. Because the Earth's gravitational field diminishes with distance, the gravitational force system on the satellite is equivalent to a force through the satellite mass center and a torque about it. This torque can be exploited to orient the satellite along the local vertical. (This torque is the same one that keeps one face of the moon pointing toward the Earth.) To create an appreciable torque, the satellite should be long and skinny, have one moment of inertia that is much smaller than the other two. The long axis is oriented along the local vertical. Ideally, the satellite should be dumbbell-shaped. The gravitational gradient torque is derived as follows[9]: From Fig. 4 we get

$$\bar{T} = \int \bar{p} \times d\bar{F} = GM_e \int \bar{p} \times \frac{\bar{R}_0}{|\bar{R}_0 + \bar{p}|^3}\, dm \tag{8}$$

If we expand the denominator in a series, using that the satellite is small compared with the orbit radius, Eq. (8) becomes

$$\bar{T} = -\frac{GM_e}{R_0^3}\int_m \left(\bar{p}\times\bar{R}_0\right)\left[1-3\frac{\bar{R}_0\cdot\bar{p}}{R_0^2}\right]\mathrm{d}m \qquad (9)$$

Since each component of the cross product is a term that is linear in a $\bar{p}$ component, the "1" inside the square bracket results in terms that vanish if we locate the satellite coordinate origin at the satellite mass center. Consequently,

$$\bar{T} = 3n^2\int_m \left(\bar{p}\times\frac{\bar{R}_0}{|R_0|}\right)\left(\frac{\bar{R}_0\cdot\bar{p}}{|R_0|}\right)\mathrm{d}m \qquad (10)$$

and

$$n \triangleq \sqrt{GM_e/R_0^3}$$

is the mean motion of the satellite at distance R_0. The important part of the torque can be derived by considering a "linear" satellite, i.e., one having (see Fig. 4) $L/D \gg 1$. The torque is "stabilizing" in the direction of $\bar{p}_R = \bar{p}\times\bar{R}_0$ and given by rearranging Eq. (10).

$$\bar{T} = (3/2)n^2 A \sin(2\alpha)\hat{p}_R$$

$$[(3/2)n^2 A] = 2.0\times10^{-3}\ \mathrm{N-m} \quad \text{(Transit)} \qquad (11)$$

wherein

$$A \triangleq \int \bar{p}^2\,\mathrm{d}m$$

is the moment of inertia about a transverse axis. A general treatment of Eq. (10) alters Eq. (11) (for a satellite symmetrical about the stabilized axis) by replacing (A) with $(A-C)$, where C is the moment of inertia about the symmetry axis. For $C<A$, the torque is a restoring one, but the satellite is indifferent as to which end is up. Consequently, care must be exercised when the orientation is initially established.

For the Transit satellites, $C/A = 0.02$ is achieved by extending a 30.5-m boom with 1.27 kg on the end. (The boom is stored for launching on a motorized reel, similar to a carpenter's steel tape.) The tape is internally stressed when fabricated so that, as it unfurls, it curls about its long axis and forms a circular cross-section boom 1.27 cm in diameter (deHaviland Corporation). The initial attitude is established by first magnetically stabilizing the satellite with an electromagnetic dipole along its symmetry axis. The magnet in the satellite is energized so that the "right" end of the satellite is up when the satellite is over the Earth's north magnetic pole. At the pole, the Earth's magnetic field lies along the local vertical; the boom is extended and the satellite magnet de-energized. The satellite is captured "right-side" up with the antenna pointing toward the Earth. It was of course not this easy. In the development phase, before retractable booms were available and before this scheme for capturing the satellite right-side up was developed, there were problems with satellites captured upside down and limited means of reorienting them. The first successfully gravity gradient stabilized satellite was Transit 5A-3 launched on June 15, 1963. The 1988 state of the art limits the rms deviation of the satellite symmetry axis from the local vertical to 1 or 2 deg.

Two subtle design problems have been solved by simulation/experience: just how long to make the boom and how heavily to damp the motion. Both depend on the level of the disturbing torques; a principal source is solar radiation pressure. In a sense, the orientation is always disturbed because the local vertical continuously changes direction as the satellite orbits the Earth. The motion must be continuously damped so

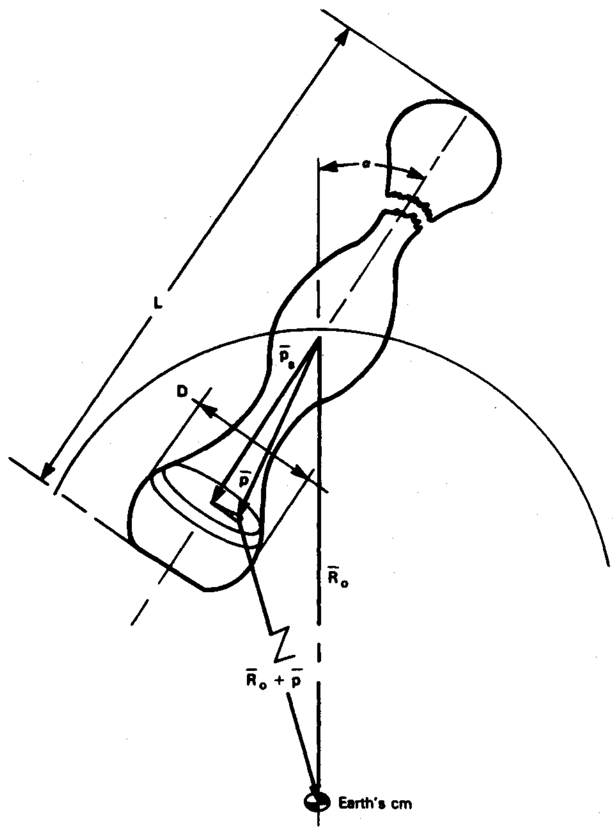


Fig. 4 Computation of gravitational gradient torque.

that the symmetry axis will continuously follow a changing direction.

Gravity gradient stabilization has an unsolved problem: After performing satisfactorily for five years, Oscar-18 became unstable in September 1974. The sun was on the normal to the orbit plane. The attitude settled down within a month and became slightly disturbed the following March when the sun was on the other end of the orbit normal. The attitude became unstable again the following September, and the boom broke. Although thermally induced boom bending plays a role, there are aspects of the problem that are not understood.

XI. Attitude Determination System

In developing the attitude control system, it was necessary to develop a system for determining the attitude. The satellite attitude is specified by a rotation matrix, which transforms a vector resolution from a satellite-fixed coordinate frame to a reference (Earth-fixed) coordinate frame. Over a year or two the following system evolved.

Aboard the satellite we simultaneously measure the direction cosines of two vectors and telemeter these six numbers to the ground. The satellite-sun line and the (vector) components of the Earth's magnetic field are two convenient vectors. Moreover, since both have known resolutions in the reference frame, they satisfy another necessary condition (see below); we can write for either vector

$$\boldsymbol{S} = \{A\}\boldsymbol{s}$$

$$\boldsymbol{M} = \{A\}\boldsymbol{m} \qquad (12)$$

where $\boldsymbol{s}$, $\boldsymbol{m}$ are the solar and magnetic (unit) vectors measured aboard the satellite and $\boldsymbol{S}$, $\boldsymbol{M}$ are their corresponding resolutions in the reference coordinate system. The $\boldsymbol{S}$ is obtained from a solar ephemeris and $\boldsymbol{M}$ from the satellite position and an analytical model of the magnetic field. $\{A\}$ is the instantaneous orthogonal rotation matrix.

We can solve Eqs. (12) for $\{A\}$ as follows: $\{A\}$ also transforms the cross product

$$\frac{S \times M}{|S \times M|} = \{A\} \frac{s \times m}{|s \times m|} \tag{13}$$

We can now concatenate the three equations into a single matrix equation.

$$\left\{ S \,\vdots\, M \,\vdots\, \frac{S \times M}{|S \times M|} \right\} = \{A\} \left\{ s \,\vdots\, m \,\vdots\, \frac{s \times m}{|s \times m|} \right\} \tag{14}$$

This can be solved for $\{A\}$ by inverting the matrix on the far right. This has an inverse as long as s and m are distinct, which is almost always the case. There is one minor problem, the measurements (the lower case letters) are noisy and the straightforward numerical inversion and subsequent multiplication will not yield an orthogonal matrix for $\{A\}$. There is an easy fix. We replace the middle column m on the right side of Eq. (14) with

$$\frac{s \times m}{|s \times m|} \times s$$

and make a similar change to the left side. After this change, the matrices on the left and on the extreme right of Eq. (14) will be orthogonal and $\{A\}$ can be produced by transposing. The $\{A\}$ will consequently be orthogonal.

This system was implemented on a number of Transit satellites. The solar sensors were individual solar cells in low impedance circuits. When a cell is illuminated, its current is proportional to the cosine of the angle between the cell's normal and the sun. The current from illuminated cells, distributed over the satellite, were combined to give the direction cosines of the sun vector. Flux gate magnetometer sensors along orthogonal axes in the satellite produced m. Consistency between scalar invariants ($s \cdot m$ and $S \cdot M$) and between ($|\bar{M}|$ and $|\bar{m}|$) were used to edit the data and to calibrate the magnetometer sensors in orbit. The system is accurate to about 1 deg, most of the error coming from errors in the model of the magnetic field.[10] With minor interpretation changes, Eq. (14) is a general geometrical theorem for determining the orientation of any rigid body.

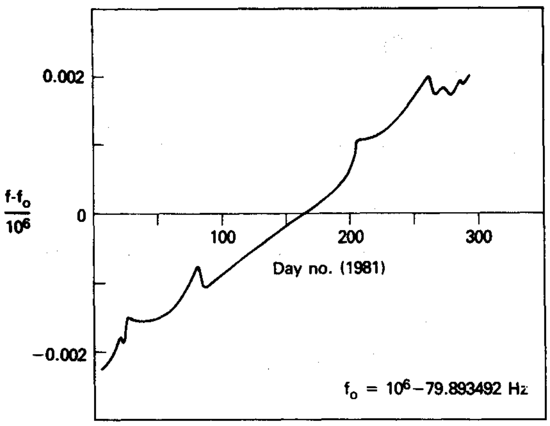


Fig. 5a Satellite oscillator frequency (Transit "Oscar" 13); the satellite was launched on May 18, 1967.

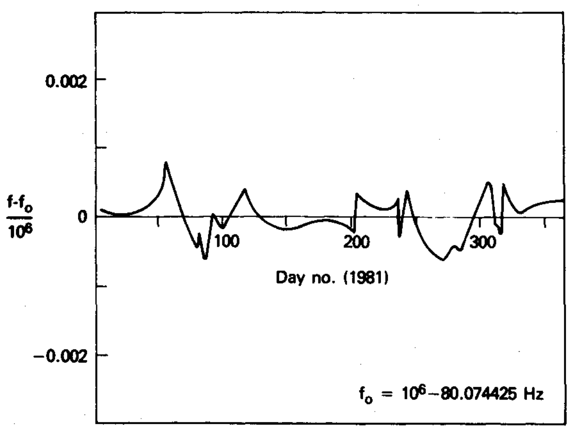


Fig. 5b Satellite oscillator frequency (Transit "Oscar" 19); the satellite was launched on August 27, 1970.

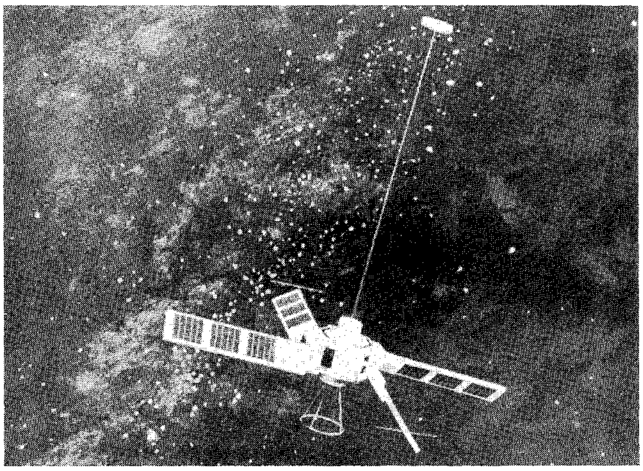


Fig. 6 The "Oscar" Transit satellite design ca. 1966. Thirty-two of these have been built and all but four launched.

XII. Satellite Oscillator

The oscillator is the "heart beat" of the satellite. It is the source, after suitable multiplication, of the radiated Doppler signals. The noise on the oscillator frequency must be as low as possible.

About 1959 W. H. Guier wrote a computer simulation of the navigator's computation and derived empirically that the navigation error component due to short-term oscillator noise has an upper bound given by

$$E = 3.6 \times 10^{11} \, \Delta f/f \quad \text{(m)} \tag{15}$$

Typically, the error is one-half this large; so the rule-of-thumb is "A frequency error of 10^{-11} gives about 2 m in navigation error." With this rule as a guide and a commitment to produce 0.1 n.mi. (185 m) navigation accuracy, there was a requirement to build oscillators having noise levels of 10^{-10} or less. Additionally, the oscillator had to satisfy stringent weight, volume, and power constraints as does any instrument in a satellite.

The requirement was first met about 1963 with a 1-kg oscillator having a volume of 1.2 liters.

Several design characteristics have received careful attention (J. R. Norton, private communication, 1988).

1) Increasingly rigorous attention has been paid to thermal control of the oscillator: a) The first oscillators were thermally isolated from the remaining spacecraft structure by attaching them with nylon lacing. b) Circuitry and crystals were enclosed in double-walled Dewar flasks insulated with alternate layers of fiberglass and aluminized plastic. (It is unnecessary to evacuate the flask; the vacuum in orbit is free.) c) The second and subsequent oscillators had increasingly sophisticated thermal control.[11]

2) Careful selection of electronic parts and careful circuit design is required to insure that a) all parts operate well within their specified bounds and b) the normally present charged-particle environment will not stress the parts.

3) Following the technology and taking advantage of higher-Q, lower noise crystals is important.

Figures 5a and 5b show one step in the evolutionary design process. The data shown in Fig. 5a is for Oscar 13, launched on May 18, 1967. (This satellite operated continuously for 21.7 years and failed in January of 1989.) Figure 5b shows the oscillator frequency for 1982 (2 points/day). The drift rate is 4×10^{-9} per year. The oscillator frequency of the more recently built Oscar-19 (launched August 22, 1970, expired December 5, 1984) is shown in Fig. 5b. There is no perceptible long-term drift here.

XIII. Satellite Power System

With one exception, all the Transit satellites have used nickel cadmium batteries and solar cells to power the satellite. They have proved remarkably reliable and durable. The solar cells are mounted on panels that are erected in orbit (see Fig. 6).

As a design strategy, these panels accomplish three things: they provide the needed flexibility to mount any number of solar cells, they uncouple the thermal design of the solar panels from the thermal/structural design of the spacecraft body, and they simplify the assembly and testing of the satellite.

The nickel cadmium batteries and the solar cells that charge them were sized using a simulation of the (orbital) period fraction that the satellite spends in the sun. After five years of development, three design rules emerged that are necessary for long satellite life (W. E. Allen, private communication, 1988).

1) Shallow depths of discharge are necessary for long battery life. Typically, no more that 10% of the rated capacity should be removed during discharge.

2) Once the battery is fully charged, to trickle charge it at a rate that exceeds one-tenth of its capacity will seriously impair its life; a fully charged 10 W-h battery should not be "force fed" at a rate that exceeds 1 W-h/h at room temperature.

3) Interconnecting the individual solar cells must be carefully done, otherwise the differential thermal expansion between the cells and the mounting panel will break the connections.

XIV. Satellite Memory Processor

We could not buy the 25k-bit memory to store and read out the satellite ephemeris. We had to design and build it. As J. Perschy (private communication, 1988) describes the memory design history:

> The early satellites utilized magnetic core diode logic for the memory system. Ring counters were fabricated utilizing magnetic cores as storage elements separated by diodes and capacitors as temporary storage during a shift pulse. The only transistors used were those furnishing the shift-pulse current for the magnetic cores. This current was steered through the switching core into a magnetic core memory plane to partially select a memory element. The entire memory system consisted of a few dozen transistors, several hundred (logic) magnetic cores, diodes, and capacitors, plus twenty five thousand memory cores and lots of copper wire. Although the detailed electrical design went smoothly, the fabrication was a real nightmare. Winding the logic cores either required long tedious hours manually winding cores and counting the windings or trying to keep a winding machine in adjustment that was not designed for the job. Stripping the #46 copper wire (diam $= 1.6 \times 10^{-3}$ in.) and attaching it to a post proved to be an exceptional challenge. Attaching by welding was tried first, then soldering. The wires would tend to break after a module was "potted" in foam for vibration resistance. Repair was not easy! Many hours were spent digging through foam looking for a broken segment of the tiny wire. Nevertheless, six of these memory systems were fabricated and three were flown during the early years. The lesson we learned was a bit surprising. The most reliable component of the memory system was the switching transistor. We had tried to minimize their number! The least reliable was the fine wire connections. To this day this lesson is valid. Transformers are avoided and transistors, in integrated circuits, are used by the millions.

XV. Fitting the Transit System into the Historical Context

Satellite navigation/surveying clearly challenged the importance of traditional techniques and as such met resistance from the practitioners of those arts. This opposition challenged us (drove us "nuts") to fit the new techniques into the traditional art so that results could be directly compared. This turned out to be difficult because the differences arose at the heart of both systems. Traditional surveying and satellite surveying have, as their underlying bases, different primitives and different "observables." From studying both, the satellite system relieves the limitations of the older arts—particularly in replacing an ellipsoid model of the mean sea level surface ("the geoid") with a consistently employed, high-precision, geopotential model; and since the satellite is near Earth, the surveyors height is readily available. The satellite-based system is highly redundant, i.e., it is capable of self-improvement using data gathered from the system; improvements made in the hardware and models can be consistently reflected throughout. Further discussion of these ideas can be found in Ref. 12.

XVI. Management of the System Development

The successful development of the system crucially depended on the leadership of R. B. Kershner (1913–1982). Kershner was a wonderfully talented man. In addition to rare technical skill and judgment, Kershner had "guts" (courage), a deep integrity, and a delightful sense of humor. He understood the difference between the craft of management and the art of leadership and the limitations of the former. As a technical leader, Kershner had *style*; a style that can glimpsed from his papers on the subject of technical "management."[13-16] He prevailed against the darkness.

Frosch[17] reflects similar attitudes on how to develop a large system.

XVII. Concluding Remarks

Unknown to us at the time, participating in the Transit program turned out to be a once-in-a-lifetime opportunity. We kept looking for the chance to participate in another program of equal or greater significance. I imagine that the people who worked on Project Apollo had similar problems.

Appendix

Doppler Effect

If an oscillator of frequency f_s (measured in a coordinate system fixed with respect to the oscillator) moves in a straight line with speed v toward an observer, then the observer measures a frequency

$$f = \frac{f_s}{1 - v/c} \tag{A1}$$

wherein c is the "phase velocity" of the wave front in the intervening medium. This equation is derived from purely classical considerations (see derivation in Ref. 18 or 19).

Expanding to first order in v/c we obtain

$$f = f_s + (f_s/c)v \tag{A2}$$

The additive term

$$\Delta f = (f_s/c)v \tag{A3}$$

is the classical, first order in v/c, Doppler shift. This expression is inadequate in several ways for describing the satellite-observer case. The first inadequacy is easy to remedy: In the preceding derivation the (satellite) oscillator directly approaches the observer. We generalize by noting Fig. A1. The previous definition of (the scalar) v is generalized and replaced with

$$v = -\frac{\mathrm{d}}{\mathrm{d}t}|\bar{r} - \bar{r}_T| \tag{A4}$$

This expression acknowledges that the Doppler shift is a consequence of the relative motion between the observer and satellite, that both can move, and that the oscillator can either approach or recede from the observer. The observer "moves," even though he might be stationary relative to the Earth, because the satellite motion is conveniently described in inertial space; in inertial space the observer rotates with the Earth. The minus sign in Eq. (A4) is consistent with the derivation of Eq. (A3): When the oscillator and observer directly approach each other, the indicated differentiation gives for this *special case*

$$\frac{\mathrm{d}}{\mathrm{d}t}|\bar{r} - \bar{r}_T| \doteq -|\dot{\bar{r}} - \dot{\bar{r}}_T| \tag{A5}$$

and the two minus signs in Eqs. (A4) and (A5) cancel. Equation (A4) is the more generally applicable form. As the satellite moves toward us, the frequency shift is positive, "toward the blue."

In the context of the classical theory, there is one other subtlety. A wave front emanating from the satellite is immune to the motion of the satellite *except* at the time when the radiation is emitted; similarly, the wave front knows nothing of the observer's motion until it arrives at the observer. This implies that we should compute the satellite position (and velocity) at times t^* and observer coordinates at times t where (t^*, t) are related by

$$t^* = t - \frac{|\bar{r}(t^*) - \bar{r}_T(t)|}{c} \tag{A6}$$

Equation (A6) is the basis of associating instants (epochs) on two clocks, one in the satellite (t^*) and another on the ground (t).

The appropriate *classical* expression for the Doppler-shifted frequency is then

$$f = f_s + \Delta f \tag{A7}$$

wherein the Doppler shift is

$$\Delta f = -\frac{f_s}{c}\frac{\mathrm{d}}{\mathrm{d}t}|\bar{r}(t^*) - \bar{r}_T(t)| \tag{A8}$$

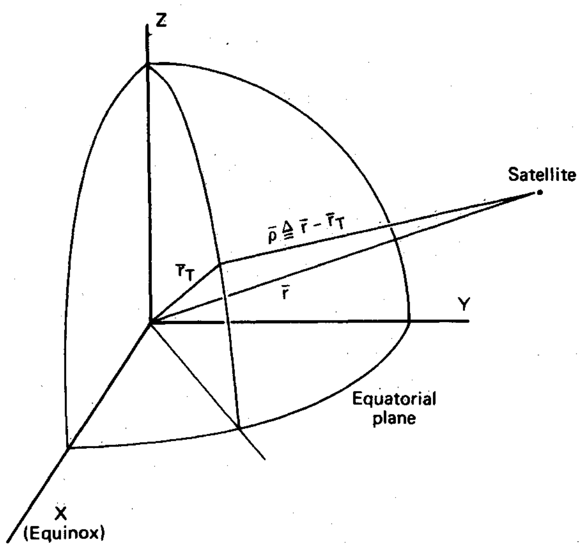


Fig. A1 Geometry of satellite pass with nomenclature.

and t^*, t are related by Eq. (A6).

Carrying out the indicated differentiation we obtain:

$$\Delta f = -\frac{f_s}{c}\hat{\rho}\cdot\left(\dot{\bar{r}} - \dot{\bar{r}}_T\right)\left(1 - \hat{\rho}\cdot\frac{\dot{\bar{r}}}{c}\right) \tag{A9}$$

wherein

$$\bar{\rho} \triangleq \bar{r}(t^*) - \bar{r}_T(t) \tag{A10}$$

$|\bar{\rho}|$ is the "slant range" (see Fig. A1); ($\hat{}$) denotes a unit vector. Something unusual has happened. From the expansion to first order in v/c [see Eq. (A2)] we obtain a second-order term. Since we know that relativistic effects add second-order terms in v/c, the expression is incomplete. The second-order term in Eq. (A9) of classical origin is, however, the largest of these terms. The development to this point is strictly classical. If we begin anew using relativity theory, Jenkins[20,21] obtains the complete expression, correct to second order in v/c

$$\frac{\Delta f}{f_s} = -\left\{\frac{1}{c}\hat{\rho}\cdot\left(\dot{\bar{r}} - \dot{\bar{r}}_T\right)\left(1 - \hat{\rho}\cdot\frac{\dot{\bar{r}}}{c}\right) + \frac{1}{2c^2}\left(\dot{\bar{r}}^2 - \dot{\bar{r}}_T^2\right)\right.$$
$$\left. - \frac{1}{c^2}\left[\frac{GM}{|\bar{r}_T|} - \frac{GM}{|\bar{r}|}\right] + 0\left(\frac{1}{c^3}\right) + \delta fa\right\} \tag{A11}$$

where $\bar{r}$, $\dot{\bar{r}}$ are the position and velocity of the satellite at the time (t^*) the signal leaves the satellite; $\bar{r}_T$, $\dot{\bar{r}}_T$ are the position and velocity of the tracking station (observer) at the time (t) the signal arrives at the station; f_s, c are the satellite transmitter frequency and the speed of light, respectively; GM is the gravitational constant times the mass of the Earth; and $\bar{\rho}$ is defined by Eq. (A10). The instants t^*, t are in turn related by Eq. (A6).

Relativity theory adds two additional terms of order $(v/c)^2$, the "time dilation" and the "gravitational red shift" terms. The final term acknowledges that there are "atmospheric" correction terms. Both the ionosphere and troposphere have measurable effects on the Doppler shift and must be considered.

Doppler's original paper is cited in Ref. 22.

Acknowledgment

I am indebted to a number of people for the content of this paper: To my fellow staff members at Johns Hopkins APL particularly, R. E. Fischell, W. H. Guier, J. R. Norton, J. W. Perschy, W. E. Allen, and J. B. Moffett who answered innumerable questions. To John Junkins of Texas A&M University who made many helpful suggestions. To Christine Hilliard for her able assistance. We are deeply indebted to the Strategic Systems Project of the U.S. Navy (Admiral Levering Smith, USN Ret.) who supported the system development, and to the Navy Astronautics Group who operate it.

Mostly, I am indebted to the more than 200 people who worked on Project Transit with intensity and dedication; because of their realization that Transit was really worth doing *and* because of the effective leadership of R. B. Kershner.

References

[1]Black, H. D., Jenkins, R. E., and Pryor, L. L., "The Transit System, 1975," Johns Hopkins Univ. Applied Physics Lab., Laurel, MD, Rept. TG 1305, 1976.

[2]"Bibliography on the Navy Navigation Satellite Program, 1958–1979," Johns Hopkins Univ. Applied Physics Lab., Laurel, MD, Rept. SR 80-3, Jan. 1, 1980, p. 203.

[3]Weiffenbach, G. C., "The Genesis of Transit," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. AES-22, No. 4, 1986, pp. 474–482.

[4]Guier, W. H., and Weiffenbach, G. C., "Theoretical Analysis of Doppler-Radio Signals from Earth Satellites," Johns Hopkins Univ. Applied Physics Lab., Laurel, MD, Rept. BB-2761, 1958; see also

Proceedings of the IRE, Vol. 48, April 1960, pp. 507-516.

[5]Kershner, R. B., "Present State of Navigation by Doppler Measurement from Near Earth Satellites," *APL Technical Digest,* Vol. 5 No. 2, 1965, pp. 2-9.

[6]Wyatt, T., "The Gestation of Transit as Perceived by One Participant," *Johns Hopkins APL Technical Digest,* Vol. 2, No. 1, 1981, pp. 32-38.

[7]Kershner, R. B., "Technical Innovations in the APL Space Department," *Johns Hopkins APL Technical Digest,* Vol. 1, No. 4, 1980, pp. 264-278.

[8]Fischell, R. E., "Magnetic Damping of the Angular Motions of Earth Satellites," *ARS Journal,* Vol. 31, No. 9, 1961, pp. 1210-1217.

[9]Newton, R. R., "Method of Stabilizing an Astronomical Telescope," *ARS Journal,* Vol. 29, No. 9, 1959, pp. 665, 666.

[10]Black, H. D., "A Passive System for Determining the Attitude of a Satellite," *AIAA Journal,* Vol. 2, No. 7, 1964, pp. 1350-1351.

[11]Norton, J. R., "An Ultrastable Low Power 5 Mhz Quartz Oscillator Qualified for Space Usage," *Proceedings of the 30th Annual Frequency Control Symposium,* Electronic Industries Association, Washington, DC, 1976, pp. 275-278.

[12]Black, H. D., "The Transit System, 1977," *Philosophical Transactions of the Royal Society of London,* Vol. A294, 1980, pp. 217-236.

[13]Kershner, R. B., "The Size of Research and Engineering Teams," *Proceedings of the 11th National Conference on the Administration of Research,* Pennsylvania State Univ. Press, University Park, PA, Sept. 1957, pp. 77-83.

[14]Kershner, R. B., "The Arcane Art of Research and Development Management," *Johns Hopkins APL Technical Digest,* Vol. 2, No.1, 1981, pp. 45-49.

[15]Kershner, R. B., "The Cost Benefit Monster," *Johns Hopkins APL Technical Digest,* Vol. 2, No. 3, 1981, pp. 207, 208.

[16]Kershner, R. B., "Where Have all the Underruns Gone?" *Johns Hopkins APL Technical Digest,* Vol. 2, No. 4, 1981, pp. 327-329.

[17]Frosch, R. A., "A New Look at Systems Engineering," *IEEE Spectrum,* Vol. 6, No. 9, 1969, pp. 24-28.

[18]Feynman, R. P., et al., *The Feynman Lectures on Physics,* Vol. 1, Addison-Wesley, Reading, MA, 1963, p. 34-7.

[19]Semat, H., *Fundamentals of Physics,* Rinehart, New York, 1951, pp. 398, 399.

[20]Jenkins, R. E., "Relativistic Effects in Doppler Tracking," Applied Physics Lab., Johns Hopkins Univ., Laurel, MD, TG 1033, Oct. 1968, p. 23.

[21]Jenkins, R. E., "A Satellite Observation of the Relativistic Doppler Shift," *Astronomical Journal,* Vol. 74, No. 7, 1969, pp. 960-963.

[22]Doppler, C., "Ueber das farbige Licht der Dopplesterne und einiger anderer Gestirne des Himmels," *Abhandlungen der Königlichen Böhmischen Gesellschaft der Wissenschaften, Fünfter Folge Zweiter Band von den Jahren 1841-1842,* Prague, 1843, pp. 467-482.