Design Studies for a Technology Assessment Receiver for The Global Positioning System

Final Report for NASA GRANT 1639

December 15, 1981

by

John H. Painter

Department of Electrical Engineering
Texas A & M University
College Station, Texas 77843

TABLE OF CONTENTS

1. Navigation Fundamentals and Orbit Characterization.

The basic mathematical ideas underlying satellite radio navigation are no different from those underlying any other kind of navigation employing independent measurements to determine position. First, a coordinate system is chosen by means of which the navigator's position is described. For the present purpose, a right-hand Cartesian System using components x,y, and z may be taken with origin the center of the Earth. Such a system is called "Earth-centered". It is usual practice to take the z-axis through the true North Pole with a positive sense. Then the x and y axes lie in the plane of the equator. Such a coordinate frame is called "Equatorial" as opposed to the "Ecliptic" coordinate frame used in astronomy and celestial navigation.

If the x-axis is projected from the center of the Earth toward a fixed point in space at an infinite distance, such as the Vernal Equinox on the celestial sphere, the coordinate frame does not rotate with the Earth. If the x-axis is projected through a point on the Earth's surface, such as the zero-latitude, zero-longitude point, then the frame is called Earth-Centered, Earth-Fixed (ECEF). We shall deal here with either type of Earth-Centered Cartesian coordinate frame, as the need arises.

Let us now define the navigator's position in the x-y-z frame as a 3-vector, or 3 x 1 column matrix, as

$$(1): \qquad \underline{P} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Let three measurements be denoted as m_1, m_2, and m_3. In a well-posed navigation problem, there exist three scalar functions, $f_1(,,)$, $f_2(,,)$, and $f_3(,,)$ such that

$$(2): \qquad \begin{aligned} m_1 &= f_1(x,y,z) \\ m_2 &= f_2(x,y,z) \\ m_3 &= f_3(x,y,z) \end{aligned}$$

Now denote the measurements and functions by 3-vectors, as

$$(3): \qquad \underline{m} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} \qquad \underline{f} = \begin{bmatrix} f_1\ (\ ) \\ f_2\ (\ ) \\ f_3\ (\ ) \end{bmatrix}$$

Then we have a compact notation for the functional relationship as

$$(4): \qquad \underline{m} = \underline{f}\ (\underline{P})$$

where $\underline{f}\ (\ )$ is a vector-valued function of vector argument.

The navigation problem is, "Given the structure of the function, $\underline{f}\ (\ )$, and the measurements, $\underline{m}$, solve for $\underline{P}$." If, for instance, the relationship between $\underline{m}$ and $\underline{P}$ were linear and invertible, then the problem would be simple. Unfortunately, most navigation problems have no inverse function for $\underline{f}$ which can be applied straight-away to solve for $\underline{P}$.

As an example, in the problem of measuring distances from the navigator's position to three known points described by the vectors $\underline{P}_1$, $\underline{P}_2$, $\underline{P}_3$, the components of the function, $\underline{f}\ (\ )$, are given by

$$(5): \qquad f_i\ (\underline{P}) = \left\| \underline{P}_i - \underline{P} \right\| \ : i=1,2,3$$

That is, the function is the square-root of the dot-product, or the Euclidian Norm. This function is non-linear and has no inverse.

In cases where the function, $\underline{f}(\ )$, has no inverse, a so-called "perturbation" solution is generally used, based on "linearizing" the navigation problem. This is done as follows.

Assume there exists a reasonably precise estimate of the navigator's position, $\underline{P}_0$. (This might be a dead-reckoning position, for instance.) From the assumed position, $\underline{P}_0$, calculate the corresponding "predicted" measurement, $\underline{M}_0$, using the known function, $\underline{f}\ (\ )$. That is,

$$(6): \qquad \underline{M}_0 = \underline{f}\ (\underline{P}_0)$$

Now, $\underline{M}_0$ is the first term of a vector Taylor series for $\underline{M}$, expanded about the "point", $\underline{P}_0$. Calculate the coefficient of the linear second term of the series, which is the Jacobian matrix, H, given by,

$$(7): \qquad H = \left. \frac{\partial \underline{f}(\ )}{\partial \underline{P}} \right|_{\underline{P} = \underline{P}_0} \quad : \text{a } 3 \times 3 \text{ matrix}$$

Then write

$$(8) \qquad \underline{M} = \underline{M}_0 + H\,(\underline{P} - \underline{P}_0)$$

Equation (2) is a two-term (linear) approximation to $\underline{M}$. If a navigation solution exists then H will be invertible with inverse, H^{-1}, and we may solve for $\underline{P}$ as

$$(9): \qquad \underline{P} = \underline{P}_0 + H^{-1}\,(\underline{M} - \underline{M}_0)$$

In a satellite-based radio navigation system, it is the function of the navigator's radio receiver to provide the measurements m_1, m_2, and m_3, comprising $\underline{M}$. In the Global Positioning System (GPS), the measurments are of distances, or ranges, from the satellites to the navigator's receiver. In GPS, the navigation computer requires knowledge of the satellite position at the time of the measurement. Thus, the system incorporates data messages from satellite to navigator in the radio transmission. These data messages contain the Keplerian orbital elements from which the satellite orbit can be computed in the navigation processor.

It is not intended to give a complete treatise on orbital mechanics here. For that, the reader is referred to one of the standard texts, such as Battin [1]. However, some familiarization is required, since contemporary satellite radio navigation receiver-processors, such as for GPS, perform satellite orbit calculations internally. Furthermore, the satellite data messages contain the so-called orbital elements for the satellite being received. Thus, a knowledge of these elements and their use is necessary to understanding of the system itself.

The form of a stable satellite orbit is an ellipse when only the satellite and Earth are considered. That is, considering the satellite and the Earth to form a "two-body problem" in celestial mechanics, the solution for the satellite trajectory is an ellipse with the Earth (center of mass) at one focus. The elliptic orbit is fundamentally described by its semi-latus rectum, eccentricity, and time of nearest approach to Earth. These three quantities are called orbital elements. To relate this elliptic orbit to the coordinate system in use requires three more elements, classically taken as the Euler angles (defined below).

Figure 1 shows several of the orbital elements which are used to describe the orbits of contemporary radio navigation satellites.

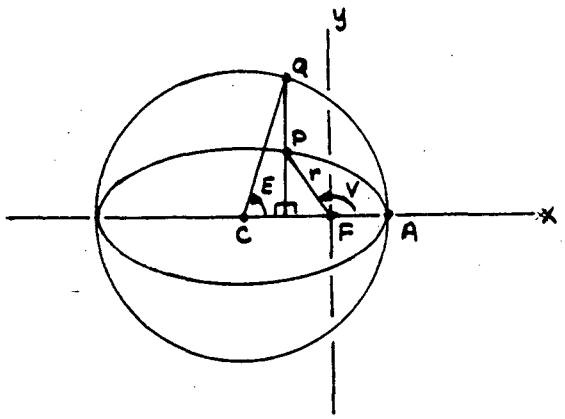

Figure 1. Elliptic Orbit Parameters.

The following quantities are identified with respect to Figure 1

$\angle$ AFP = V	:	True Anomaly
$\angle$ ACQ = E	:	Eccentric Anomaly
$\overline{AC}$ = a	:	Semi-major axis
$\overline{FP}$ = r	:	Orbit radial distance
o<e<1	:	Orbit eccentricity
P	:	Satellite position on orbit
A	:	Perigee point
F	:	Ellipse Focal Point (Position of Earth center of mass)
C	:	Center of circumscribing circle
	:	Earth universal gravitational parameter
$n_o = 2\pi / T = \sqrt{\mu/a^3}$	:	Mean angular motion
T	:	Orbital period
t_{oe}	:	Time of perigee passage
M	:	Mean anomaly

The orbit radial distance, $r=\overline{FP}$, is given by

(10): $$r = a (1 - e \cos E)$$

Let x" and y" denote a local Cartesian coordinate system in the plane of the ellipse with the x"-axis passing through the focus and perigee point. The y"-axis is orthogonal to x" and passes through the focus. The coordinates of the point P on the elliptic orbit are then given by

(11): $$x" = a (\cos E - e) = r \cos V$$
$$y" = a (1-e^2)^{1/2} \sin E = r \sin V$$

The eccentric anomaly, E, is determined by solving Kepler's equation, given as

(12): $$n(t-t_{oe}) = M = E - e \sin E$$

In (12), the quantity, t, is the time at which the value of E (and x" and y") are to be determined. The values of t and t_{oe} must be measured with respect to the same clock.

Equations (11) and (12) define the orbit postion as a function of time, referenced to the orbital plane, major axis, and perigee time. The equations define a right-hand, Earth-Centered Cartesian coordinate system, x"-y"-z", for which z" is identically zero. Now, there are two other right-hand, Earth-Centered, Cartesian systems which are of interest and in which the orbit must be defined.

Both of the systems are known as equatorial, since in both, the x and y axes lie in the plane of the equator. The z axes both pass through the true North Pole. The first coordinate system is called inertial (relative only to the Earth), since the x-axis is directed at a fixed point in the firmament. That point is the First Point of Aries, or Vernal Equinox. It is the point at which the sun rises above the plane of the equator. The second coordinate system is called Earth-Fixed, since the x-axis passes through the Greenwich meridian of Longitude. (0^{o}).

In a two-body problem with an Earth of sperically symmetric mass density, and no air friction or powered flight, the elliptic orbit is fixed relative to the inertial coordinate system. Thus if we denote the inertial system as x'-y'-z', these new coordinates may be obtained from the x"-y"-z" by three successive simple (one-axis) rotations about the origin (the center of the Earth).

To aid in visualization of the required rotations, Figure 2 is given below.

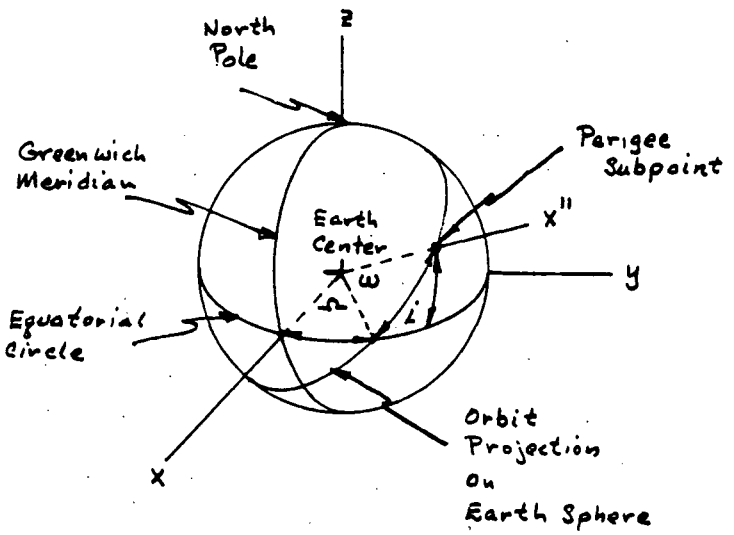

Figure 2. Orbit Coordinate Conversion.

In the figure, the orbital plane is shown inclined to the equatorial plane by an Inclination angle, i. Also, the orbital major axis, which passes through the perigee point, is displaced in the orbital plane from the intersection point with the equatorial plane by an angle, ω, called the Argument of Perigee. The intersection point between the two planes, with the satellite z coordinate increasing, is called the Ascending Node Crossing. The intersection of the two planes, called the Line of Nodes, is angularly displaced Eastward of the x-axis in the equatorial plane by an amount Ω. This latter angle is called the Right Ascension of the Ascending Node. These three angles are the Euler angles.

From the above view of the angular relationships between the elliptic orbit in x", y", z" and the Earth-Centered Inertial frame in x', y', z', it is clear what rotations are necessary to transform from the former to the latter. The rotation order is ω, i, and Ω. The compound transformation is given in vector-matrix form as

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \Lambda \cdot \begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix}$$

(13):

$$\Lambda = \begin{bmatrix} \cos\omega\cos\Omega - \sin\omega\cos i\sin\Omega & -\sin\omega\cos\Omega - \cos\omega\cos i\sin\Omega & \sin i\sin\Omega \\ \cos\omega\sin\Omega + \sin\omega\cos i\cos\Omega & -\sin\omega\sin\Omega + \cos\omega\cos i\cos\Omega & -\sin i\cos\Omega \\ \sin\omega\sin i & \cos\omega\sin i & \cos i \end{bmatrix}$$

The transformation, Λ, brings the satellite into Earth-Centered-Intertial coordinate system, x', y', z', by a rotation about the Earth's rotational axis by an angle, Ω, the Right Ascension of the Ascending Node. To bring the position into the ECEF system, x,y,z, where x projects through the Greenwich meridian at the equator can be done by redefining Ω, and using the transformation, Λ.

Let Ω now be redefined to be the longitude of the Ascending Node, measured positive westward from the Greenwich meridian. The angle Ω is now time-varying, due to the rotation of the Earth and due to the precession of the orbital plane in inertial space. Thus, define

(14): $$\Omega = (\Omega_0 - \dot{\Omega}_e \cdot t_{oe}) + (\dot{\Omega} - \dot{\Omega}_e)\cdot(t - t_{oe})$$

In (14), the first parenthetical expression gives Ω_0, the Longitude of the Ascending Node at t=0, corrected by the easterly rotation rate of the Earth, $\dot{\Omega}_e$, acting up to t=toe. The second parenthetical expression then corrects Ω to the time, t, for Earth rotation and for precession of the orbital plane at the rate $\dot{\Omega}$, positive westerly. The parameter, t_{oe}, is called "epoch time" and is aligned with time of perigee passage.

Because a two-body problem is not sufficiently accurate, and the Earth's gravitational field is not spherically symmetric, and other reasons, the parameters of the orbit are not stable. Thus, a host of real-time corrections is needed for precise orbit calculations. In particular, the Mean Motion, n,

Inclination, i, Right Ascension of Ascending Node, Ω, and Mean Anomaly, M, are all specified at t_{oe} and corrected at subsequent times using rate factors. Also, the "Argument of Latitude", $\phi = v + \omega$, as well as satellite orbit radius, r, and inclination, i, are all corrected for the second zonal harmonics of the gravitational field.

The order of computations, with corrections, is given below in Table 1.

$\mu = 3.986008 \times 10^{14}$: WGS-72 gravitational parameter, m^3/s^2

$\dot{\Omega}_e = 7.292115147 \times 10^{-5}$: : WGS-72 Earth rotation rate rad/sec

$C_{rc}, C_{rs}, C_{uc}, C_{us}, C_{ic}, C_{is}$: Zonal harmonic coefficients, given

t_{oe} : time of epoch, given

e : eccentricity, given

a : semi-major axis, given

$n_o = (\mu/A^3)1/2$

$n = n_o + \Delta n$: Δn given

$M = M_o + n(t - t_{oe})$: M_o given

$M = E - e\sin E$: Solved implicitly for E

$\cos v = (\cos E - e)/(1 - e\cos E)$

$\sin v = (1 - e^2)^{1/2} \sin E/(1 - e\cos E)$

$v = \arctan(\sin v/\cos v)$

$\phi = v + \omega$: ω given

$u = \phi + \delta u$; $\delta u = C_{uc} \cdot \cos 2\phi + C_{us} \cdot \sin 2\phi$

$r = a(1 - e\cos E) + \delta r$; $\delta_r = C_{rc} \cdot \cos 2\phi + C_{rs} \cdot \sin 2\phi$

$x'' = r \cos u$

$y'' = r \sin u$

$i = i_o + \delta_i = C_{ic} \cdot \cos 2\phi + C_{is} \cdot \sin 2\phi$, i_o given

$\Omega = (\Omega_o - \dot{\Omega}_e \cdot t_{oe}) + (\dot{\Omega} - \dot{\Omega}_e) \cdot (t - t_{oe})$; $\Omega_o, \dot{\Omega}$ given

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \Lambda(\omega, \Omega, i) \cdot \begin{bmatrix} x'' \\ y'' \\ o \end{bmatrix}; \Lambda(\cdot) \text{ given by (13)}$$

Table 1. Satellite Orbit Computations.

2. Corrections for Ionospheric Propagation.

The ionosphere is the interface between our atmosphere and outer space. It is a region of ionized atmospheric gases and free electrons which have not recombined with the atmospheric ions whence they came. Historically, the ionosphere was described as being composed of "layers" which were thought to give rise to "reflection" of radio waves from the Earth. Now, it is known that the ionosphere is described by a density of electrons per cubic centimeter which has a rapid onset in the neighborhood of 50 kilometers altitude, has a maximum density in the neighborhood of 250-450 kilometers, and decreases more gradually to zero in the neighborhood of 800-3000 kilometers.

Satellite radio-navigation frequencies are sufficiently high so that there are no marked refractions of the ray paths or absorption. However, because of the accuracies desired, the small refractive effects are significant and must be accounted for in precision navigation.

The significant effects of ionospheric passage upon the radio wave are two. First, the wave propagates at an infinitesmally smaller velocity in the ionosphere than in vacuum. Second, as the wave propagates through a region of changing density the wave is refracted smoothly, according to the gradient of the density. This means that from a point above the ionosphere to a point below the ionosphere, the wave travels via a curved path which is longer than the straight path connecting the two points. This causes deleterious effects on a ranging system, such as GPS.

Because the ray-path from satellite to user is not straight, the range measured by a ranging system is always a little long. Since the curved ray path always lies above the straight-line path, and therefore sees a greater projection component of satellite velocity, the Doppler frequency measured by a Doppler system is always a little greater in magnitude. However, this is a very small effect compared to the effect upon a Doppler system of just the excess propagation delay of the curved path. The effects upon GPS will be explained in more detail below.

The bending of the ray path is a function of refractive index, μ, or rather of its gradient along the path. μ is a function of frequency, F in Hz, and electron density per cubic centimeter, N, given as

(15): $\mu(N) = [1-80.6N/F^2]^{1/2}$

When μ becomes imaginary, the gradient becomes infinite and a radio wave cannot penetrate the ionosphere further. However, for satellite navigation frequencies and relatively large peak electron densities, μ is very near 1 and small bending of the ray path occurs. For a value of frequency of 1575MHz, μ is of the order of

(16): $\mu_{1575} = 1 - \pi/2 \times 10^{-11}$

Likewise, the angular separation of the straight and curved paths is very small. For a satellite to user range of 25,000 km. and an excess delay of 20 meters, the path separation angle at the satellite is of the order of 6 minutes of arc.

Graphs of results are easily plotted with satellite elevation angle, γ, as abscissa and excess range, ΔR, as ordinate. Plots may be made for fixed Solar Zenith Angle, χ, or Local Hour Angle, ϕ, with the other serving as parameter for a family of plots. Figure (3) shows an example for two values of N_{max} for a local hour angle of $+45^{\circ}$, which is roughtly 15:00 hours, local time. A family of curves for solar zenith angle, χ, from 0° to 90° is shown. [8]

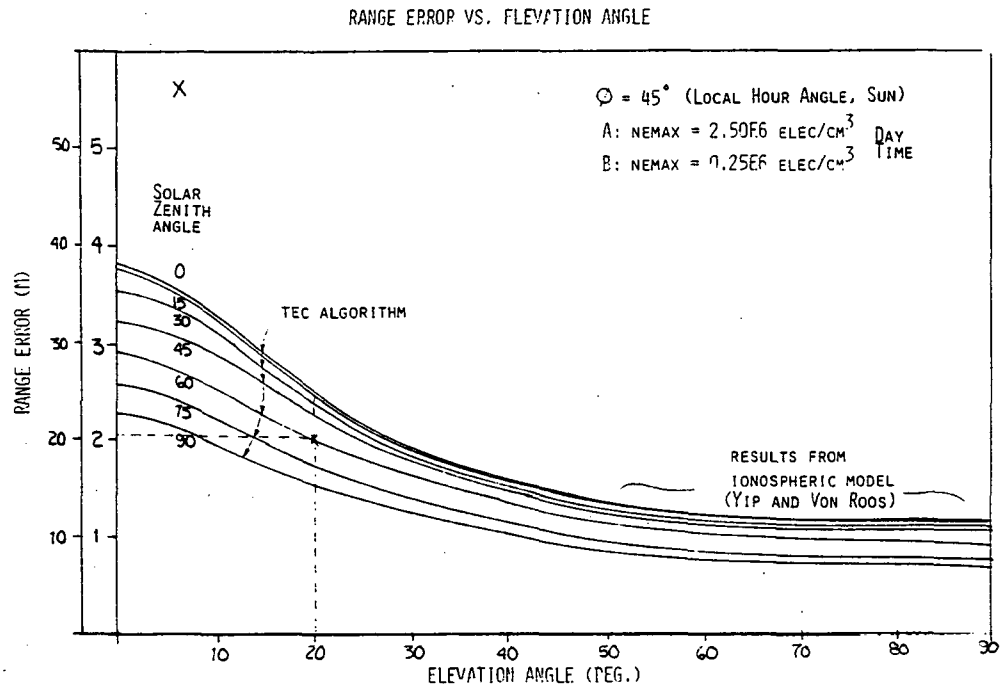

Figure 3. Ionospheric Range Error.

An interesting feature of the example above is that the excess delay does not increase markedly for lower elevation angles as has been held previously. The increase from 90^0 to 0^0 elevation angle is seen to be only by a factor of 2-3. The absolute value of R is seen to be linearly dependent on N_{max}.

In GPS the effect of R is to bias the measured position away from the satellite. If three satellites of apprximately the same elevation angle, but separated in azimuth by 120^0, can be quickly received, then the R biases tend to cancel in the Latitude-Longitude coordinates. The effect on altitude remains, however, with the tendency being to measure lower than the true altitude. For surface navigation, this is not a problem.

With respect to correcting the effect, there are several possibilities. The first is to use two separate transmission frequencies to correct the excess delay effects. This technique takes advantage of the fact that excess delay is inversely proportional to the square of frequency.

Suppose that two frequencies f_1, and f_2 are used for transmission, with $f_1<f_2$ and with the frequencies phase-coherently generated from the same oscillator source so that

(17): $f_2/f_1 = (K)^{1/2}$: K known, stable

Now, denote the ranges measured at frequencies f_1, f_2, as, respectively

(18): $R_1 = R + \Delta R_1$
$R_2 = R + \Delta R_2$

where R is true range and R_1, R_2 are the excess ranges at frequencies f_1 and f_2. There is a constant C_1 such that

(19): $\Delta R_1 = C_1/(f_1)^2$, $R_2 = C_1/(f_2)^2$

Thus,

(20): $\Delta R_1 = (f_2/f_1)^2 \Delta R_2$

and

(21): $R_1-R_2 = \Delta R_1-\Delta R_2$

From these equations it follows that

$R = [R_2(f_2/f_1)^2 - R_1]/[(f_2/f_1)^2 - 1] =$

(22): $= (KR_2 - R_1)/(K - 1)$

where K is given by (17)

Thus, from (22) the true range R may be determined by measuring apparent ranges R_1 and R_2 at two known frequencies, f_1 and f_2, whose ratio is a stable constant, K.

In the case of Doppler, let the Doppler at frequency, f, be denoted by D, where

(23): $D = V_R/C \cdot f$

and V_R is radial velocity component (signed) with C the speed of light. Taking time derivatives of (23), we obtain

$$\dot{R}f_1/C = [D_2(f_2/f_1) - D_1]/[(f_2/f_1)^2 - 1] =$$

(24): $$= [\sqrt{K}\, D_2 - D_1]/(K - 1)$$

The quantity $(\dot{R}/C)f_1$ is true Doppler at frequency f_1. D_2 and D_1 are the apparent Dopplers measured at frequencies f_2 and f_1 respectively.

Equations (22) and (24) show the two-frequency method for correcting Ranging or Doppler measurements for the effects of ionospherically-induced excess range.

3. GPS System and Signal Structure.

The NAVSTAR Global Positioning System, or GPS, is a system employing, ultimately, eighteen satellites in 12-hour orbits of 55 degrees inclination. The system is being implemented by the Department of Defense for military use. However, it has a "Clear Access" C/A channel which is available for general civil use. The GPS development program grew from two 1960's programs, the Air Force's 621-B program and the Navy's Timation program which were merged in 1973. GPS is a second-generation satellite navigation system which applies the pseudo-noise (PN) ranging technology developed by NASA [9] in the 60's to the satellite navigation technology embodied in Transit.

The GPS employs satellites which are precisely controlled in their orbital positions. Indeed, knowledge of a set of orbital elements, or ephemerides, over one year old is sufficient to predict satellite visibility times within five minutes from a known Earth position. The GPS is a passive navigation system on the part of the user, in that only reception of satellite-transmitted signals is used by the navigator to compute position. GPS uses simultaneous, or near-simultaneous reception of signals from four satellites to compute three coordinates of position and one of time difference, due to error between satellite and user clocks. Thus, given visability of four satellites, GPS provides the capability for continuous navigation processing, rather than isolated position fixes interconnected with dead reckoning. As will be described below, positions may be computed on the order of every second. Hence, GPS provides the capability to "track" vehicles characterized by high dynamic maneuvering.

The positions of the GPS satellites are uniformly distributed about the Earth, three per orbit, in six orbits.[1] Thus, with the full orbital constellation of eighteen satellites, more than four are normally visible at any given time, anywhere on Earth. This leads to a problem discussed below of selecting the four satellites having the "strongest" navigation geometry. Also,

---

[1]According to plans revealed by USAF Space Division as of the time of this writing.

if a user vehicle should lose line of sight with one GPS satellite, because, say, of blockage of signal by vehicle superstructure, then the possibility exists of immediately "acquiring" another satellite to replace the one just lost. This procedure is the same as if one of four satellites dropped below the horizon. The latter occurence can be predicted, of course, and acquisition of a replacement satellite planned ahead.

The GPS navigation-processor operates from measurements of range between satellite and user. Actually, the measurement is one of elapsed time between the time of satellite transmission of a known signal reference and the time of user reception of that same reference. Given the speed of light, the elapsed time measurement equates to a measured range. Now, the satellite transmission time is measured according to its clock, which is precisely set to GPS System Time, with an error of the order of 3 nano-seconds. However, the user reception time is measured with respect to the user clock, based on a user reference frequency oscillator. The user oscillator is generally of much lower quality than the satellite frequency standards by many order of magnitude. The GPS demostration satellites, the so-called NDS-series, have carried both the Cesium and Rubidium frequency standards with basic long term frequency stabilities of the order of 10^{-12} to 10^{-13} ($\Delta f/fo$). This is truly amazing stability when one realizes that the frequency offset just due to relativity effects is of the order of 4 parts in 10^{-10} [10]. Thus, the satellites are able to maintain a System reference time to within several nano-seconds over a 12 hour period, which equates to about 1 meter of range. The user frequency reference oscillators (for a supposed reasonable cost civil user) have long-term frequency stabilities of the order of 10^{-7} to 10^{-9}. This stability yields clock drifts of 1 to 100 nano-seconds per second, equivalent to 0.3 to 30 meters of range per second. Thus, it is necessary for the inclusion of a fourth coordinate in the user's position, this coordinate being user clock bias (in meters).

Because the measured range is always in error by the clock bias, the measured quantity is called pseudo-range. Pseudo-ranges without correction are processed directly, as detailed below, to solve for the four user coordinates.

GPS operates with two available satellite signal frequencies, for the purpose of removing the unknown ionospheric additional delay time. These frequencies are 1575.420 MHz, called L1, and 1227.600 MHz, called L2. These two

Carrier wave frequencies are phase-coherently generated by frequency multiplication of the same basic standard frequency of 10.23 MHz. The multiplication factors are 154 for L1 and 120 for L2. Actually, the satellite basic reference is lowered from 10.23 by $4.45 \times 10^{-10} \times 10.23$ to equalize the relatively effect, but this is not significant to our consideration of signal structure.

The L1 transmitted carrier frequency carries two signals in phase quadrature. One phase carries the C/A ranging signal and a 50-baud data-link. The orthogonal phase carries a Precision PN ranging signal (the P-code) and the same 50-baud data-link. The exact phase relationship is

$$\text{C/A Carrier Phase} = \text{P Carrier Phase} + 90^{o}$$

The L2 transmitted carrier may carry either the C/A or the P-code.

The ranging signals and data-link signal are digital, of varying baud rates. The P-code rate is 1.023 Mega-baud. The data-link rate is 50-baud. The data-link bits are combined with the code bits using modulo-two addition (the exclusive-or). The composite code-data bit stream is then modulated onto the proper carrier phase using full binary phase-shift-keying (+ or - 90^{o}), which leaves no unmodulated carrier residual in the transmitted spectrum. That is, no carrier phase reference is transmitted. Neither are any clock references transmitted for the various digital signals. Thus, it is to be expected that some of the chief problems in receiving these signals have to do with achieving signal synchronization.

The PN ranging codes are special examples of Maximal-Length, Linear, Shift-Register Sequences [11] , called Gold codes, after their inventor, Robert Gold. The structures of the particular GPS Gold codes are explained in detail in [12]. The unambiguous length, or repetition period, of the C/A codes is 1023 chips, or 1.0 milli-second, exactly. The basic time period is also related to the data-bit clock period, since there are exactly 20 C/A code epochs per data bit. Since the C/A code is only 1.0 milli-second long, it can resolve ranges unambiguously only within multiples of about 300 kilometers. The resolution of the ambiguity is performed during the initial satellite acquisition process as a part of obtaining the first "fix".

There is really only one P-code, which is the product of two PN codes whose lengths in chips are relatively prime numbers (that is, with no comman divisors). These two codes are of lengths 15,345,000 and 15,345,037 chips, respectively. Since a chip period is 100 nano-seconds, the length of the product code is more than 38 weeks. P-codes for individual satellites are taken as non-overlapping one-week segments of the 38-week-long code. The long code is re-started at midnight Saturday-Sunday UMT (Greenwich), every week.

The digital data, transmitted by each satellite contains all the Keplerian orbital parameters for the satellite position computations shown in Table 1. Additionally, reduced accuracy orbital elements are contained in one satellite's data message for rough computation of the other satellites' positions. The precise orbital information is called "Ephemeris" while the less precise data is called "Almanac".

The satellite data transmision uses 30-bit words, with 10 words per each 6-second sub-frame. 5 sub-frames completes one satellite data message. With 20 milli-seconds per bit, each word is 0.6 seconds in duration, each sub-frame takes 6 seconds, and a complete message requires 30 seconds. Every 30-second data message contains Almanac for one of the 18 possible satellites. Thus, to acquire Almanac for all satellites requires 9 minutes.

Each 6-second subframe begins with a 30-bit Telemetry Word (TLM). The first 8-bits of TLM word is the hexadecimal "8B". Due to the way the data is detected in a receiver, the bits may be complemented. The final two bits of the TLM word should be zeros. If they are ones, the data is inverted. Thus, a search for "8B" or its complement establishes sub-frame synchronization.

The second 30-bit word in each sub-frame is a Hand-over Word (HOW). This word is the number of 1.5-second epochs (X1-epochs) which will have occurred since the beginning of the GPS week, at the beginning of the next sub-frame (TLM word). The purpose of this word is to enable rough synchronization of the P-code (for the first time) at the beginning of the next sub-frame. The HOW-word is an acquition aid for P-code.

The structure of the 30-bit data message is indicated in Table 4, below.

Subframe	Structure		
1	TLM	HOW	DATA BLOCK I - CLOCK CORRECTION
2	TLM	HOW	DATA BLOCK II - EPHEMERIS
3	TLM	HOW	DATA BLOCK II - EPHEMERIS, cont'd
4	TLM	HOW	MESSAGE BLOCK
5	TLM	HOW	DATA BLOCK III - ALMANAC

Table 4. Data Message Structure

Subframe 1 contains Data Block I, which contains four parameters for making a quadratic correction to the indicated Satellite clock time. Also in DB-I are eight parameters for making a rough correction for ionospheric delay for those users not equipped with dual-frequency (L1-L2) receivers. Data Block II occupies both sub-frames 2 and 3 and contains the complete, accurate ephemeris for the satellite being received. Sub-frame 4 makes provision for special broadcast messages. Sub-frame 5 contains the rotating alamanacs for all satellites.

Of interest, besides the structure of the transmitted signals, is the signal to noise ratio environment in which the signals will be received. In designing or analysing the performance of satellite to Earth links, it is usual to formulate the ratio of received signal power divided by the value of the white noise power spectral density effective in the receiver. This is the so-called C/N_0 ratio.

The received signal power, C, is equal to that transmitted, multiplied by various gain and loss factors which affect the link. Chief among these is the "space loss", which is just the attenuation of power in the transmitted electro-magnetic wave due to spherical speading of the wave front with distance from the source. This loss factor, L_s, is given by

(25): $$L_s = [V_c/(4\pi FR)]^2$$

In (25) V_c is velocity of light, R is range, and F is frequency in Hertz. The units for V_c and R must be compatible, i.e., meters/second and meters.

It is seen that the space loss varies inversely as the square of frequency and of the range. Thus, $0<L_s<1$, and L_s will always diminish the received power over that transmitted. Since L_s varies with range, there is a maximum and minimum value for L_s, depending on the elevation angle of the satellite as seen from the user's position, assumed near the surface of the Earth. The maximum range (at 0^0 elevation angle) and minimum range, respectively, are

$$R_{max} = 25{,}231 \text{ Km.}$$
$$R_{min} = 19{,}652 \text{ Km.}$$

assuming a spherical Earth of mean radius 6,371 Km. Thus, we have maximum and minimum values for L_s at frequencies L1 and L2 as given in Table 5. In Table 5, both the scalar value and the decibel value of the losses are given, where XdB = $10 \log_{10}(x)$.

	L1	L2
R_{min}	5.9460×10^{-19}	9.7926×10^{-19}
	-182.3 dB	-180.1 dB
R_{max}	3.607×10^{-19}	5.9408×10^{-19}
	-184.4 dB	-182.3 dB

Table 5. L1/L2 Loss Factors

The amount of noise effective in the receiver is, for GPS, essentially the noise generated in the radio-frequency preamplifier which is of special low-noise design. This assumes that the low-noise preamplifier has sufficient gain (20-30 dB) to over-ride the noise generated in the first stages of the receiver, itself. In this case, the white noise spectral density, N_0, is given by

(26): $N_0 = K\ T_S$

Where T_S is the "system noise temperature" in degrees Kelvin. T_S is the noise temperature of the preamplifier, itself, plus any contributions from cable and connector losses between antenna and preamplifier. Great pains are usually taken to mount the preamplifier as near to the antenna as possible. The constant K is Boltzmann's Constant given by

(27): $K = 1.38 \times 10^{-23}$ watts/Hz/^{o}Kelvin

The system noise temperature, T_S, may be specified according to the standard "noise figure", F, for the system by

(28): $T_S = 290\ (F-1)$

A handy device for determining the values of available C/N_0 for the various cases is a Design Control Table. Such a table enters the various parameters of the link such as transmitted powers, gains, losses, noise spectral density, etc., in order to arrive at a value for available C/N_0. This value is then used to calculate performance levels of various parts of the receiver. An example table for the C/A channel at L1 is given below.

The following comments are made with respect to Table 6. In parameter number 2., Modulation Loss accounts for the fact that the total satellite transmitter power is apportioned between C/A and P signals at L1 and the signal at L2. The power proportions are

(29): C/A (L1): P(L1): L2 = 4: 2: 1

In the Table, tolerances account for uncertainties in specifications, changes with age or temperature, or variations in the geometric relation between user and satellite. For a user with an antenna which is nominally omni-directional over the upper hemisphere, there is a large change in gain between the zenith direction and the horizon. The tolerance reflects this. Also, an omni-directional antenna which is circularly polarized at the zenith becomes elliptical at lower elevation angles. Hence, the tolerance for polarization loss.

Parameter	Nom. dB	Tol., dB
1. Transmitter Power	13.0 dBW	0
2. Modulation Loss, Pwr. Split	- 2.4	0
3. Circuit Loss, xmt.	- 0.3	-0.1
4. Antenna Gain, xmt.	+13.7	-2.2
5. Antenna Pointing Loss, xmt	0	0
6. Space Loss		
F=1575.4 MHz.		
R=19,652 KM	-182.3	-2.1
7. Antenna Polarization Loss	0	-1.4
8. Antenna Gain, Rcv.(Omni)	+ 3.0	-5.0
9. Antenna Pointing Loss, Rcv.	0	0
10. Circuit Loss, Rcv.	- 0.3	-0.2
11. Net Link Loss (Add 2thru10)	-168.6	-11.0
12. Total Received Power, C (1+11)	-155.6 dBW	-11.0
13. System Noise Spectral Density, N_0		
$+1.0$		
Preamp N.F. 3.0dB -0.0		
Loss Tempr. Inc. $+20^{o}K$, nom.	-204.0 dBW/Hz.	$\pm$ 1.9
14. Received C/N_0 (12-13)	+48.4 dB	-12.9 dB

Table 6. Design Control Table, L1, C/A.

The nominal value for C/N_0 in the example is 48.4 dB, +0., -12.9 dB. Thus, the value available is +48.4 dB in the best case and becomes, perhaps, as small as +35.5 dB in the worst case. Experience has shown that the adverse tolerances hardly ever add linearly, except for those which reflect correlated effects. Three such correlated tolerances are those for space loss, Antenna Polarization Loss, and Antenna Gain for receiving. The sum of these three correlated tolerances is -8.5 dB., which can be expected to occur at the low elevation angles, below, say, 10^{o}.

The availability of 48.4 dB for C/N_0 in the best case sets the upper limit for the receiver processing possibilities in the example shown. Using a rule of thumb that 10 dB or greater signal to noise ratios are required for good

processing indicates that the widest final processing bandwidth can be no greater than about 7 KHz. Being more conservative, in view of the adverse tolerances, might bring this final processing bandwidth down to, say, 3 KHz. The actual bandwidth used for final processing will depend, of course, on more detailed considerations. But, at least, the value of available C/N_o sets the scale for further thought.

4. PN Ranging for GPS.

The basic idea behind PN ranging is the following. There are two identical PN codes involved. One code propagates from satellite to user. The other code is maintained in the user's equipment. Since a PN code is completely deterministic, the sequence of one's and zero's in the code is completely known. Both codes are started up in synchronism at a particular known time, t_o. The code which propagates between satellite and user is received by the user, delayed by the propagation time. The user then shifts his own local version of the code, keeping track of the amount of time delay injected into the "local code", until its sequence just exactly matches the sequence being received from the satellite. When the user observes that the two versions of the same PN code were "matched" or synchronized at the observation time, t_r, he notes that the range-delay time, ΔT, at time t_r, was just that amount of time-delay which the user injected into the local code to cause it to match the received code.

The matching of the local code with the received code is essentially matching the zero-one sequences and then "exactly" matching the leading edges and trailing edges of each bit (or chip as they are called in ranging) within the sequences. Any error in matching the codes in time delay translates into an error in determining the range separating the satellite and user. The conversion between time delay error, ΔT, and range error, ΔR, using a velocity of light, $V_c = 2.99739 \times 10^8$ m/sec. is

(30): $\Delta R/\Delta T = 0.983$ ft/nano-second

In order to make the code matching problem yield very precise time measurements, the time duration of each code chip is made very small. For the C/A code, the chip duration is 0.9775 micro-seconds. For P-code, the duration is 97.75 nano-seconds. Because of the method used for matching code chip to code chip, the resolution in the delay time is even finer than the chip duration.

The use of sub-micro-second code chip durations means that the modulated bandwidth of the ranging signal is of the order of Mega-Hertz. For the case of + or - 90^o Phase-Shift-Keying, which is used in GPS to place the code modulation on the radio-frequency carrier, the modulated signal may be written in the form

(31): $s(t) = A\, c(t) \cdot \cos(2\pi f_0 t)$

In (31), A is signal amplitude, f_0 is carrier frequency in Hertz, and $c(t)$ is the analog code waveform, which is either +1 or -1 in value. That is, a digital 0-chip gives an analog +1 and a digital 1-chip gives an analog -1. Since $c(t)$ is a characteristically rectangular waveform, the frequency power spectrum has the characteristic form,

(32): $S(f) \sim \sin^2(\pi/R(f-fo))/(\pi/R(f-fo))^2$

where f is frequency in Hertz, f_0 is carrier frequency in Hertz, and R is code chip rate in chips per second. The graph of (40) is given below as Figure 4.

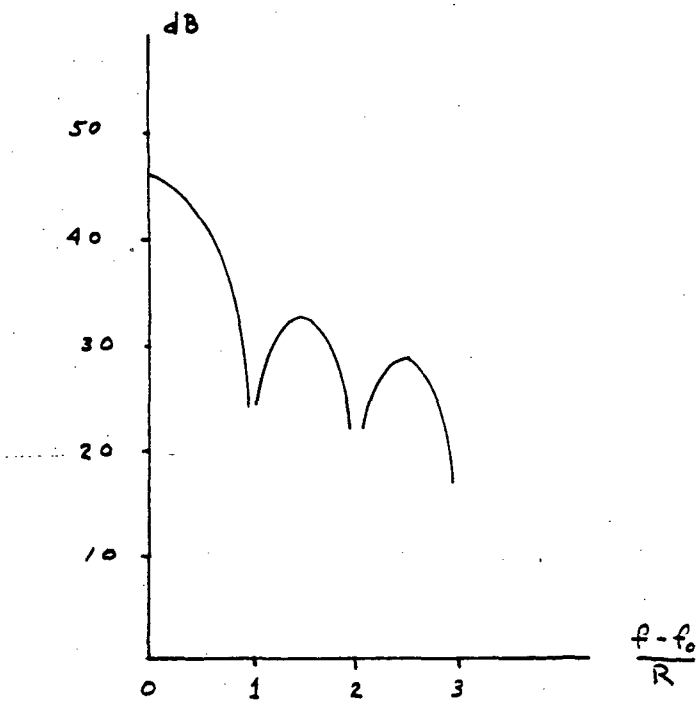

Figure 4. PN Signal Spectrum.

It is known that 92 per cent of the power in the PN signal spectrum resides between the first nulls on either side of the carrier frequency. Therefore, most receivers employ bandwidths of 2R or greater to pass the modulated signal. Now, with an available C/N_0 of 48.4 dB/Hz., the signal to noise ratio for the C/A signal in a bandwidth of 2R = 2.046 MHz. is -14.7 dB. Thus, the signal

chips are not observable in such a bandwidth, due to the over-riding effect of the accompanying noise.

Because of the impossiblity of synchronizing the received and local codes on a bit-by-bit basis, a more indirect method is used, which can operate in a smaller bandwidth. This method uses the "correlation" properties of PN codes.

Suppose we ignore for the time-being the fact that the received PN code exists as modulation upon a sinsoidal carrier waveform. Let us consider the simpler problem of synchronizing two identical PN code waveforms which exist in the + or - analog format. Let us assume that the directly received code is input to one port of an analog multiplier as C(t). Let the local code be input to the multiplier's second port as $C(t+\tau)$, denoting a slight desynchronization of amount τ seconds. Let the output of the analog multiplier be processed by a time-averager, such as a low-pass filter. This operation is shown in Figure 5.

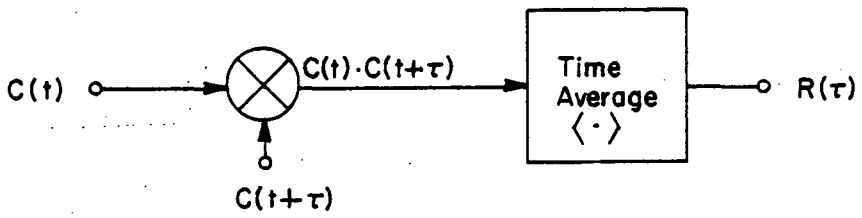

Figure 5. Cross-Correlator.

The device depicted in Figure 5 is called a "Time-Average Cross-Correlator." Let us now examine its operation on two relatively delayed versions of the same PN code. For this, let us also view Figure 6. In Figure 6 are shown several representative chips of the direct PN code and local PN code when they are near synchronism. The upper two graphs are the codes themselves, while the lower drawing is of the code product, $c(t+\tau)\cdot c(t)$.

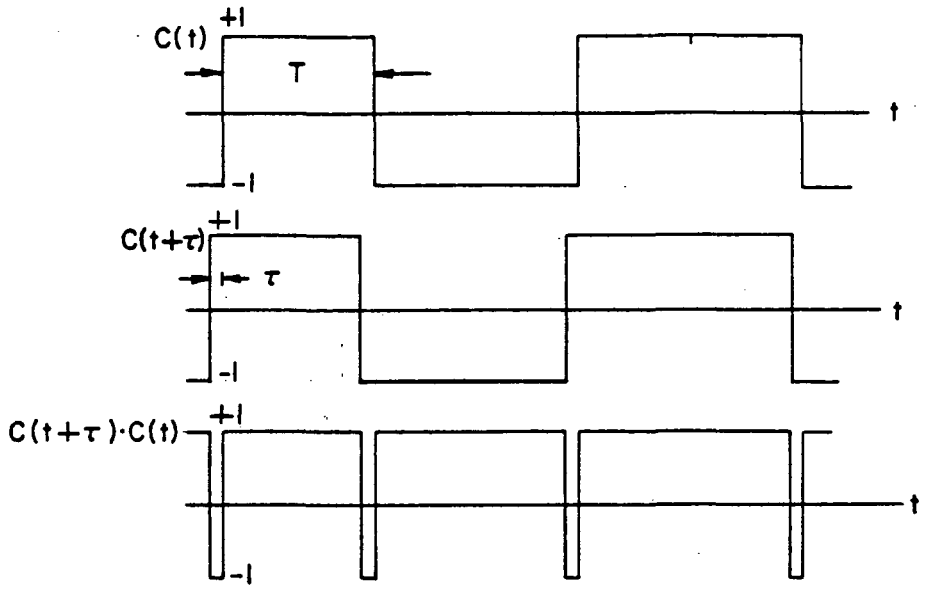

Figure 6. Relatively Delayed PN Codes.

It is clear by inspection of Figure 6 that the product, $C(t+\tau)\cdot c(t)$, is +1 most of the time, with quick excursions to the -1 state during those short periods of duration, τ, when C(t) and $C(t+\tau)$ are of opposite sign. It is also clear that the time average of the product is positive and nearly +1. When $\tau=0$, or the codes are exactly synchronized, the average is exactly +1. It can be shown that the output, $R(\tau)$, of the averager varies linearly with the offset, or relative delay, τ. [13]. Also, for a very long PN code, the average is essentially zero when C(t) and $C(t+\tau)$ are desynchronized by more than one chip period, T. This function, $R(\tau)$, which is the autocorrelation function of the basic PN code, itself, is shown in Figure 7.

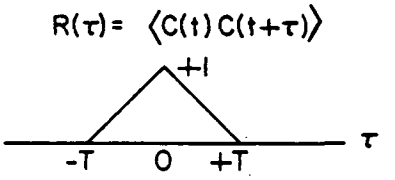

Figure 7. PN Autocorrelation Function.

The correlation properties of PN codes allow the synchronization of the received and local PN codes to be accomplished using a smaller bandwidth than that of the code itself. The lesser bandwidth is that of the time-averaging low-pass filter, employed in the cross-correlator. For example, if we wished to observe $R(\tau)$ with only + or - 5 percent noise with an available C/N_0 of 48.4 dB, a signal to noise ratio out of the averager of 26 dB would be required. This would, in turn, require the bandwidth of the averaging low pass filter to be 174 Hz.

Following on with the above example, one next asks the following interesting question. If $R(\tau) = 1$ signifies code synchronization, and it is known that the output of the correlator has + or - 5 percent noise, what is the possible ranging error incurred in accepting any single measurement of $R(\tau)$ within a + or - 5 percent neighborhood of +1? From Figure 7, with T=293 meters, equivalent, it can be easily determined that the above ranging stategy might incur an error of + or - 29.3 meters.

We have simplified the above treatment of PN code correlation to maintain visibility of the essential results. In practice, the received PN signal may exist as PSK modulation on a sinusoidal carrier, as in equation (39). The multiplication may be times the local code as PSK modulation on a sinusoid of different frequency. In this case the multiplier acts as a mixer to produce the code product $C(t)\cdot c(t+\tau)$ existing as PSK modulation upon some intermediate frequency sinusoid. The i.f. band-pass filter then does the averaging to form $R(\tau)$ which is now present in the amplitude of the output i.f. sinusoid. There

are many such ways to perform the correlation.

5. Receiver Processing Techniques

In forming $R(\tau)$ and searching out the local code delay corresponding to the center of the triangle of Figure 7, there are several possibilities. We may seek to implement a feed-back servo device in the receiver which will automatically adjust the local code delay so as to "track" the $R(\tau) = 1$ point. Otherwise, we may choose to sweep the local code past the synchronization point with the received code, observe $R(\tau)$ during the sweep, and determine, after the sweep, the local code delay which yielded $R(\tau) = 1$.

Let us consider now the received signal as it might exist at the output of a hypothetical code correlator which is implemented by a mixer in the formation of the last intermediate frequency in the receiver. Figure 8 shows the circuitry.

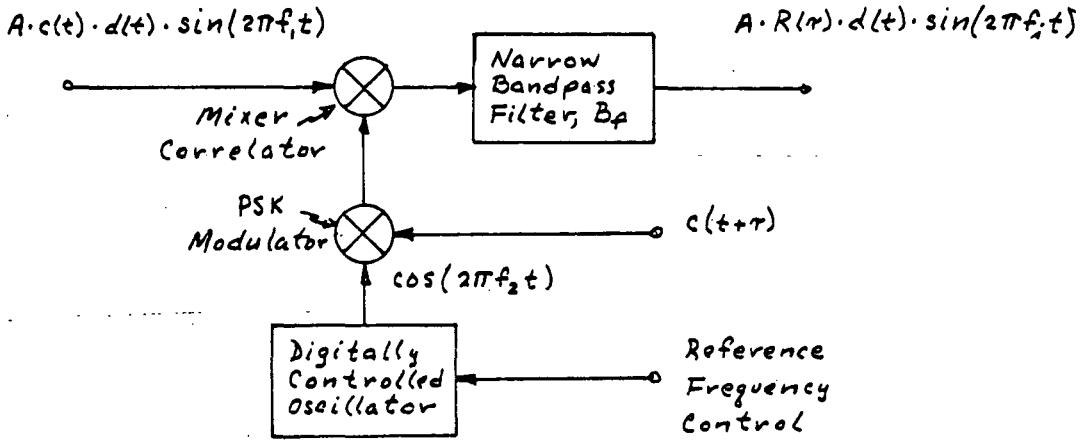

Figure 8. I.F. Correlator.

In Figure 8, the input signal to the mixer-correlator has frequency f_1. A digitally controlled oscillator produces a cosine of frequency f_2, which is phase-shift-keyed by the local code, $C(t+\tau)$. The mixer forms the code product and the bandpass filter averages it while also selecting the i.f. frequency, $f_i = f_1 - f_2$. The bandpass filter bandwidth, B_f, is small enough to average the code product, but large enough to pass the data modulation, $d(t)$.

The output of the correlator may now be processed in one of several ways. If it is desired to recover the data, then the i.f. signal may be applied to the input of a Costas Loop [16]. See Figure 9. The Costas Loop regenerates a reference sinusoid which is phase-coherent with the suppressed i.f. frequency, f_i. The reference is applied to a synchronous amplitude detector (product detector) which is also driven by the received signal. A 90-degree phase-shifted version is applied to a similar detector. The two detectors demodulate the i.f. signal and produce two signal components called the I-component (in-phase) and Q-component (quadrature phase). These components are given as

$$I(t) = A/2\ R(\tau)\ d(t)\ \cos\phi$$

(33):
$$Q(T) = -A/2\ R(\tau)\ d(t)\ \sin\phi$$

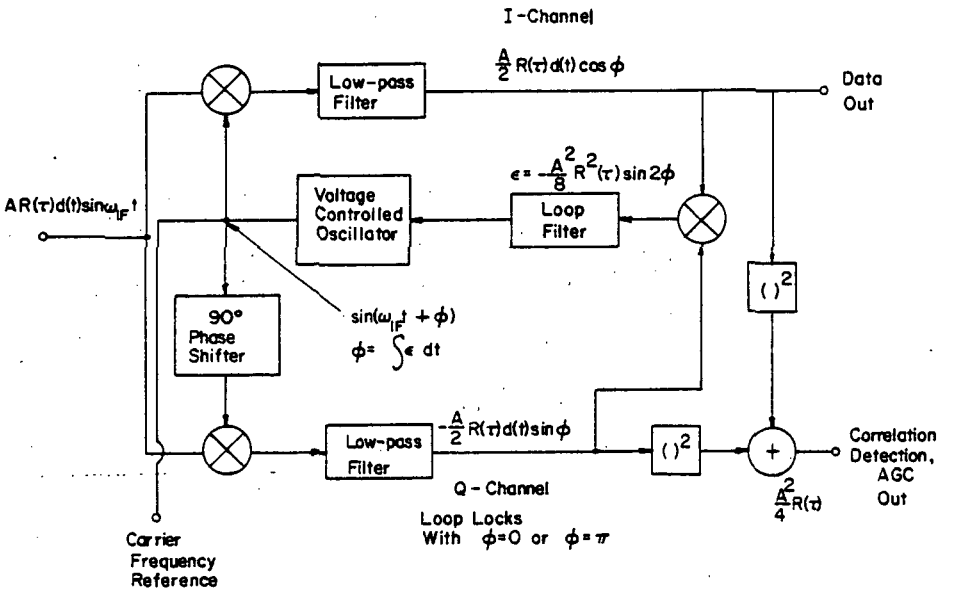

Figure 9. Costas Loop.

In equations (33), ϕ is the error (tracking) in phase between the regenerated i.f. sinusoid and the received i.f. sinusoid. For proper operation of the loop $\phi = 0$ or π and the I-channel contains the data waveform, weighted in amplitude by + or - $R(\tau)$. Note that if $\phi = \pi$, then the data in the I-channel is inverted. Taking the absolute value of the I-channel waveform just gives $A/2\ |R(\tau)| = A/2\ R(\tau)$ since $d(t) = +$ or -1. Thus, the I-channel may be used to derive both data and $R(\tau)$.

A Costas Loop synchronizes and tracks the phase (and frequency) of the i.f. signal sinusoid, and produces an output proportional to $R(\tau)$. Two Costas Loops may be employed in a feed-back loop configuration to synchronize and track the clock frequency of the received code. Such a configuration is called an "Early-Late Clock Loop". This scheme employs two correlation-mixers and two local code generators, as shown in Figure 10.

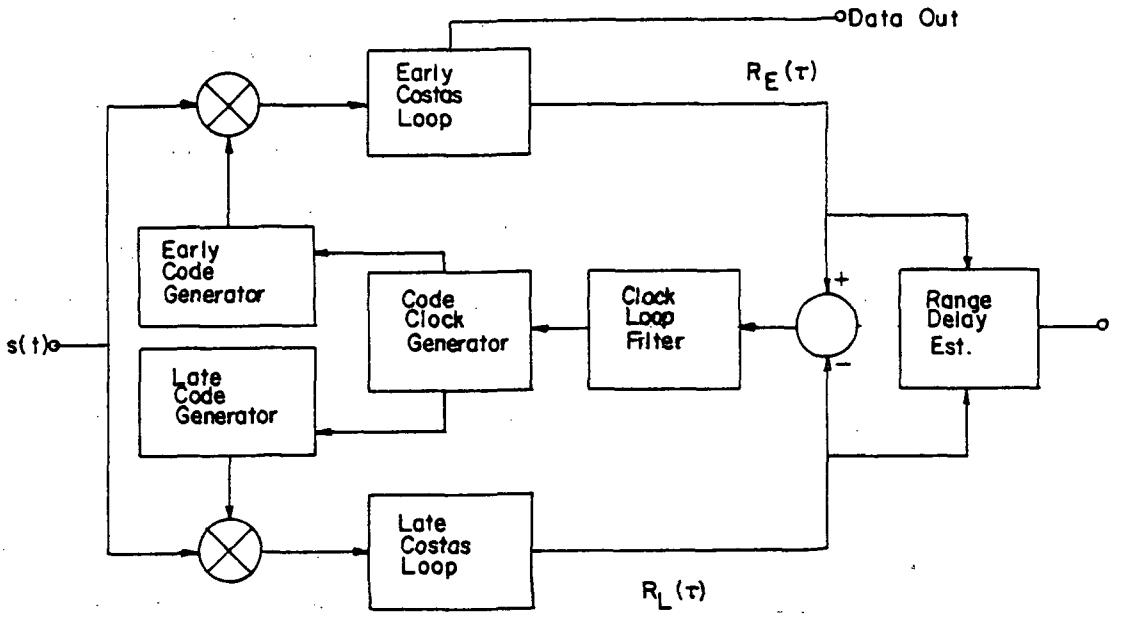

Figure 10. Early-Late Code Synchronizer.

The Early-Late Loop works in the following way. The two local code generators are driven by the same Code Clock Generator (oscillator). However, one Code Generator is advanced slightly in delay with respect to the other. Thus, the code from one generator is slightly "early" with respect to the code from the other generator (say, one bit early, for example). As both code generators are brought near the synchronization point, the "early code" passes the synchronization point and the "early Costas Loop" locks, producing an "early $R(\tau)$" called $R_E(\tau)$. When the early code is 1/2-bit past the synchronization point, the "late code" is 1/2-bit before the synchronization point, the "late Costas Loop" comes into lock, and produces a "late $R(\tau)$" called $R_L(\tau)$.

Referring again to Figure 10, we see that $R_L(\tau)$ is subtracted from $R_E(\tau)$, forming a "tracking error function" of τ. This function is shown in Figure 11. When the early and late codes are exactly 1/2 bit early and late, with respect

to the received code, then the tracking error signal is exactly zero. When the received code tends to move one way or the other from the bracketed position, the tracking error signal increases or decreases from zero and is used to adjust the frequency of the Code Clock Generator in the proper direction, to maintain "code lock". When the loop is stably locked, neither the early or late code is exactly synchronized, or "prompt". Each is out of synchronization by 1/2-bit, but the error is exactly known so the $R(\tau)$ =1 delay can be inferred.

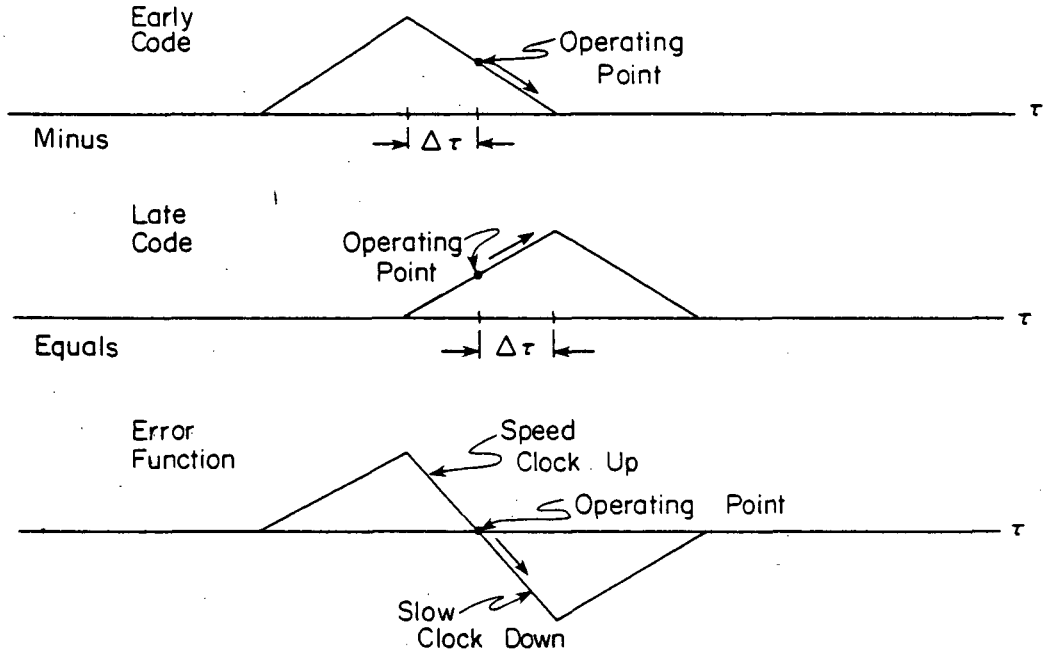

Figure 11. Tracking Error Signal.

Another processing method can be used when data is not required. This might be the case, for instance, in a two-channel, sequential receiver, where one-channel can be devoted to getting data and the other to ranging only. In the ranging only case, a pair of I and Q detectors, as in the Costas Loop, can be used with a sinusoidal reference oscillator which is not locked to the received carrier phase. If this oscillator is different in frequency by an amount Δ, from the i.f. frequency, then the I and Q channel signals are

$$I(t) = A/2\ R(\tau)\ d(t) \cos(2\pi\Delta t)$$

(34):
$$Q(t) = A/2\ R(\tau)\ d(t) \sin(2\pi\Delta t)$$

The two signals in (34) may be squared and summed to produce a signal

(35): $I^2(t) + Q^2(t) = A^2/4\ R^2(\tau)$

As the single local code generator is swept past the synchronization point, the signal $R^2(\tau)$ may be used to estimate the $R(\tau) = 1$ delay, since $0 \leq R(\tau)$ and a one-to-one relationship exists between $R(\tau)$ and $R^2(\tau)$. Since this latter processing method does not depend on having a phase or frequency lock on the i.f. sinusoid, the method is called "non-coherent". The Costas-Loop method is called "coherent".

Because of Doppler effects on the received signal, both the coherent and non-coherent ranging methods require a search, not just in delay, τ, but also in frequency. Referring back to Figure 7, the Digitally Controlled Oscillator must be adjusted to offset the Doppler effects. It is necessary that the i.f. signal be accurately centered in the narrowest processing bandwidth encountered in the remainder of the receiver. In acquiring the first satellite signal, there may be a considerable frequency uncertainty range to be searched. During normal navigation the uncertainty should be minimal.

For either coherent or non-coherent range processing, the Digitally Controlled Oscillator is preset and then the delay uncertainty region is searched. If no evidence of correlation is observed, then the DCO is reset at a different frequency, and delay searched again. In this manner the two-dimensional frequency-delay uncertainty region is searched.

Aircraft, Sequential, L1-C/A only, Incoherent Ranging

Doppler Search -	
Velocity Uncertainty	3m/s.
Doppler Search Range (+ or - $2\sigma_v$)	63
Oscillator Instability (5×10^{-7} @ fo=10.23 MHz.)	312 Hz.
Total Frequency Uncertainty	375
Range Filter Bandwidth	3000 Hz.
Delay Search -	
Position Uncertainty, Spherical	48.6m.
Delay Search Window (+ or - $3\sigma_p$)	293m.
Code Step Interval/Samples per Sweep	19.5m./16
Total Sweeps/Total Samples	32/512
Samples Dwell Time	1 ms.
Satellite Ranging Time	0.512 s.
Fix Time (4 satellites)	2.048 s.
Range Estimator SNR -	
Available C/N_0 (Best Case)	48.4 dB/Hz.
Required Bandwidth	34.8 dB Hz.
Available Ranging SNR	13.6 dB
Ranging Accuracy	10.8 m.
Error Budget -	
Range Estimator	10.8 m.
GDOP-NAV Filter Contribution (GDOP=2.0)	5.8 m.
Ionosphere and Troposphere Contribution	11.0 m.
Satellite Ephemeris Contribution	11.0 m.
Total Position Uncertainty	48.6 m.

Table 7. Design Example: Sequential, C/A Only, Incoherent Processing.

A hypothetical example of a system design is given in Table 7, to illustrate some of the above points. This design is for an aircraft user, with access only to the C/A code, on frequency, L1, employing an incoherent (phase) ranging technique. First, some assumptions are made about the accuracies of velocity and position provided by the navigation filter. The standard

deviations of velocity error and position error are taken as 3 meters/second σ_v, and 48.6 meters, σ_p, respectively. From σ_v is calculated the resulting uncertainty in received frequency due to user velocity error. It is assumed that navigation is nominal and that satellite velocity computations contribute no uncertainty. The search range due to Doppler uncertainty is taken, conservatively as + or - $2\sigma_v$, or 63 Hz. Next, the frequency uncertainty at the i.f. due to user's oscillator instability is computed as 312 Hz. The total frequency uncertainty is thus, 375 Hz. Theoretically, the bandwidth for estimating the $R(\tau)=1$ delay, using the incoherent technique need only be 375 Hz. However, since a sampled-data digital processor is used, the bandwidth is taken as the much greater value of 3 KHz., to allow fast sampling. Three times the reciprocal of this bandwidth is just the period of one epoch of the C/A code.

Next, the delay search window is taken as 293 meters, or one chip width, which is + or - $3\sigma_p$ for the position uncertainty. The $R(\tau)$ signal is repeatedly swept 32 times, collecting 16 samples per sweep, for a total of 512 samples of $R(\tau)$. The delay resolution is 19.5 meters, but the large number of samples taken reduces the quantization error to negligible proportions. With a dwell time of 1 milli-second per sample, the time to compute one satellite range measurement is 0.512 seconds. One complete sequence of four satellites takes 2.048 seconds.

Next, the performance of the ranging estimator is determined. The best available C/N_0 is taken from the previous example as 48.4 dB/Hz. In a 3 KHz, bandwidth, the resulting signal to noise ratio is 13.6 dB. The resulting estimation accuracy is determined by Monte Carlo simulation of the range processing algorithms to be σ_r=10.8 meters.

Finally an error budget is developed to compute the position uncertainty, σ_p, previously assumed. The performance of the navigation algorithms adds 5.8 meters to the range estimator error of 10.8 meters. Daytime ionospheric delay is taken at an additional 20 meter error in final position. Tropospheric error is taken as 1.0 meter. Worst case satellite ephemeris error of 11.0 meters is taken. The sum total position uncertainty is then 48.6 meters.

It is seen from the above hypothetical example that a design case is an iterative process. Since some of the design performance can only be determined by computer simulation, the design must be "fine-tuned" until all entries are consistent. The design is obviously optimized for one particular value of C/N_0.

6. The Navigation Solution.

The purpose of the navigation processor is to accept each pseudo-range measurement as it is obtained and to apply it to the solution for the user's coordinate vector, consisting of x, y, z, and b (clock bias) in the ECEF system. In so doing, the processor must solve the Keplerian equations to determine the position vector for each satellite being received. The processor also determines which satellites are visible and, of those, which four to choose for best navigation geometry.

Referring back to equations (2) through (5), the measurements in GPS are pseudo-ranges, R_i for i=1,2,3,4. We have

$$(36): \qquad R_i = f_i(x,y,z,b) = \sqrt{(x_i-x)^2+(y_i-y)^2+(z_i-z)^2} + b$$

where the ith satellite coordinates are x_i, y_i, z_i, the user coordinates are x,y,z, and user clock bias in meters is b.

As in (6) the problem is linearized by assuming a user position and clock bias $\hat{x},\hat{y},\hat{z},\hat{b}$, approximately equal to the true quantities, and the corresponding, R_i, as calculated from (36). Then, we make the first order linear expansion as

$$R_i =$$

$$= \hat{R}_i + \partial f_i(\ )/\partial x\big]_{\hat{x}} \cdot (x-\hat{x}) + \partial f_i(\ )/\partial y\big]_{\hat{y}} \cdot (y-\hat{y}) +$$

$$(37): \qquad + \partial f_i(\ )/\partial z\big]_{\hat{z}} \cdot (z-\hat{z}) + \partial f_i(\ )/\partial b\big]_{\hat{b}} \cdot (b-\hat{b})$$

We may write the equations for four satellites in vector matrix-form as

$$(38): \quad \begin{bmatrix} R_1-\hat{R}_1 \\ R_2-\hat{R}_2 \\ R_3-\hat{R}_3 \\ R_4-\hat{R}_4 \end{bmatrix} = H \cdot \begin{bmatrix} x-\hat{x} \\ y-\hat{y} \\ z-\hat{z} \\ b-\hat{b} \end{bmatrix}$$

where H is the square Jacobian matrix of ordered partial derivatives as given in (45). Now, for example, we have

$$\partial f_i(\ )/\partial X = (x-x_i)/\sqrt{(x_i-x)^2 + (y_i-y)^2 + (z_i-z)^2}$$

$$(39): \quad \partial f_i(\ )/\partial b = 1 \quad ; \quad i = 1,2,3,4$$

and so the elements of the matrix H are direction cosines for the directed distance from user position to satellite position. See Figure (12). Let α,β,γ be the angles of $\vec{R}_i$, the directed distance, with respect to the x,y,z axes, respectively. Then,

$$(40): \quad H = \begin{bmatrix} \cos\alpha_1 & \cos\beta_1 & \cos\gamma_1 & 1 \\ \cos\alpha_2 & \cos\beta_2 & \cos\gamma_2 & 1 \\ \cos\alpha_3 & \cos\beta_3 & \cos\gamma_3 & 1 \\ \cos\alpha_4 & \cos\beta_4 & \cos\gamma_4 & 1 \end{bmatrix}$$

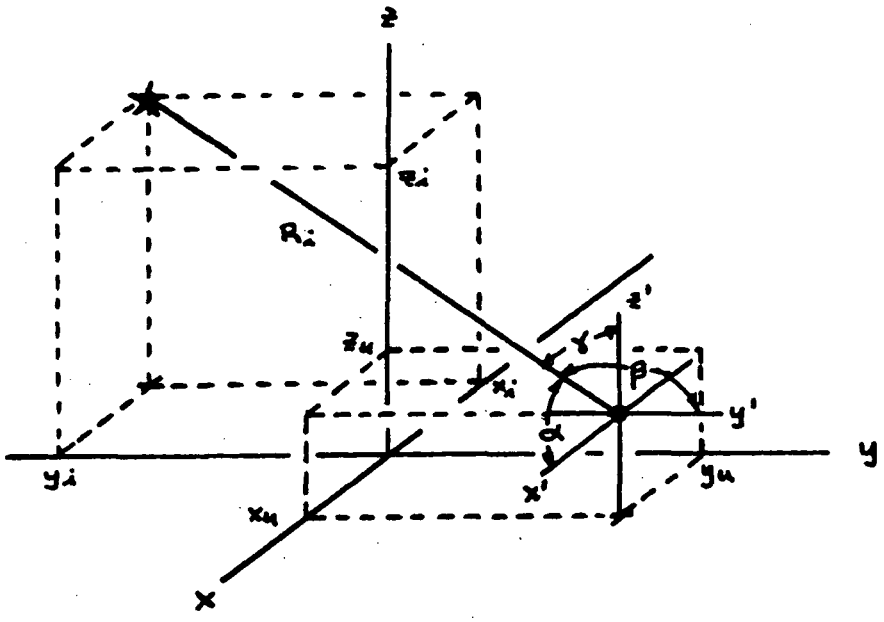

Figure 12. Relative Geometry.

Observing the structure of the H-matrix, we can see two cases in which H would not be invertible. One would be if the user was co-linear with two satellites. The other is if the 4 satellites were all on the surface of a cone.

For the nominal case, when H is invertible, we may solve for the user's coordinate "vector" as

$$(41): \quad \begin{bmatrix} x \\ y \\ z \\ b \end{bmatrix} = \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \\ \hat{b} \end{bmatrix} + H^{-1} \left\{ \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \end{bmatrix} - \begin{bmatrix} \hat{R}_1 \\ \hat{R}_2 \\ \hat{R}_3 \\ \hat{R}_4 \end{bmatrix} \right\}$$

We may define position and range vectors and their differential forms, formally as

$$(42): \quad X = \begin{bmatrix} x \\ y \\ z \\ b \end{bmatrix}, \quad \underline{R} = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \end{bmatrix} : \quad \delta\underline{x} = \underline{x} - \hat{\underline{x}}$$

$$\delta\underline{R} = \underline{R} - \hat{\underline{R}}$$

Then from (41) we have

$$(43): \quad \delta\underline{x} = H^{-1}\,\delta\underline{R}$$

which is the differential form of the linearized solution in vector-matrix notation.

Suppose, now, that we assume that the differential range vector, $\delta\underline{R}$, is subject to some random statistical error (noise). Let us assme that each element of $\delta\underline{R}$ is subject to a random noise of standard deviation, $\sigma_{\delta R}$, and that the four noises are statistically independent. The validity of this assumption depends on the mechanisms for generating the pseudo-ranges, R_i, and their linearizing estimates, $\hat{R}_i$.

Under the above assumptions we may form the covariance matrix for the differential solution, $\delta\underline{x}$, as

$$(44): \quad \text{var}\,\delta\underline{x} = \text{var}\,H^{-1}\delta\underline{R} = (\sigma_{\delta R})^2\,[H^T H]^{-1}$$

Now, let us define the standard deviation of the differential solution, as

$$(45): \sigma_{\delta p} = (\sigma_{\delta x})^2 + (\sigma_{\delta y})^2 + (\sigma_{\delta z})^2 + (\sigma_{\delta b})^2$$

The elements $(\sigma_{\delta x})^2$, $(\sigma_{\delta y})^2$, $(\sigma_{\delta z})^2$, and $(\sigma_{\delta b})^2$ are the main diagonal elements of the covariance matrix. Thus, we have

$$(46): \quad \sigma_{\delta p} = \sigma_{\delta R} \cdot \sqrt{\text{Trace}\,[H^T H]^{-1}}$$

where Trace [ ] is the matrix operator which sums the main diagonal elements.

The quantity under the diagonal has an important name in GPS. It is called "GDOP" for Geometric Dilution of Precision". It is the quantity which relates error in pseudo-range to error in 4-coordinate state solution, through the geometry as expressed in the H-matrix. We have

(47): $$\sigma_{\delta p} = \text{GDOP} \cdot \sigma_{\delta R}$$

The quantity, GDOP, includes the error in the time-bias element, b. A navigator may be more interested in just the 3-dimensional position error, without regard to b. This is the so-called PDOP, which is obtained as

$$\sigma(\delta x, \delta y, \delta z) = \text{PDOP} \cdot \sigma_{\delta R} \quad ;$$

(48): $$\text{PDOP} = \sqrt{\text{Trace } D[H^T H]^{-1} D} \quad ;$$

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The GDOP factor is used in selecting the four best satellites for the navigation problem, from those which are in view. The optimum navigation performance is obtained for the minimum value of GDOP which is always positive. The theoretical minimum GDOP occurs for four satellites each separated from all others by an angle of 120°. This value is 1.62. Unfortunately, for an Earth-surface navigator, three of the optimum satellites would be below the horizon, from the navigator's point of view.

Surprisingly enough, adding the constraint of satellite minumum visibility does not increase GDOP much. For the case of three satellites with zero-degree elevation angles, separated in azimuth by 120°, and a fourth overhead satellite, GDOP=1.74.

7. The Adaptive, Extended Kalman Filter.

A navigational filter is very useful for a vehicle which spends much of its time in uniform motion (unaccelerated). In this case a filter may be devised which reduces the effects of noise on the pseudo-range measurement, almost without limit. However, when the vehicle is subject to acceleration and higher-order motion, as in a turn, then the situation is not as good. The filter output will not necessarily track a non-uniform motion input without error. An "optimum" filter is designed to "split the difference" between errors due to measurement noise and errors due to filter mis-tracking of non-uniform vehicle motion. That is, an optimum filter design attempts to equalize some measure of the errors due to the two different sources. The measure usually is the statistical average of the squared total error.

A popular approach to the filtering problem is to describe the vehicle motion as a sample function from some family of random functions. The ranging noise is similarly described. In the case where both vehicle motion and noise are describeable as Gaussian random functions, there exists a linear filter which minimizes, over all such functions, the statistical average of the square of the instantaneous total tracking error. This filter is variously called a Kalman filter or Wiener filter. The Kalman filter is optimum from the instant it is activated. The Wiener Filter, which is historically older, is optimum only in the steady-state, after the "turn-on transient" has died away.

The starting place for modeling the class of inputs is to define the "state-vector" to be tracked. Heretofore, we have defined the "position" vector of the user to be composed of the elements x, y, z, and b. But, this is not sufficient for modelling purposes. In order to generate a higher order input model which will result in a higher order filter, we will also define time derivatives of the position components. In GPS receivers it is common practice to model position, velocity, and acceleration for x, y, z, and to model b and $\dot{b}$ for clock bias. The reason for limiting the model for b just to its first derivative is that b results from a clock oscillator whose frequency is inaccurate. There exists little acceleration in b. The same is not true for x, y, and z.

The order in which the various elements are inserted into the 11-state vector is not arbitrary and has computational consequences. However, one particular order lends itself more easily to explanation. The two popular definitions of the state vector are

$$\underline{x}^T = [x,y,z,\dot{x},\dot{y},\dot{z},\ddot{x},\ddot{y},\ddot{z},b,\dot{b}] \qquad a).$$

(49): $$\underline{x}^T = [x,\dot{x},\ddot{x},y,\dot{y},\ddot{y},z,\dot{z},\ddot{z},b,\dot{b}] \qquad b).$$

The vector-matrix generating model for x, y, z, b is, for either state-vector definition, given canonically as

$$\dot{\underline{x}}(t) = A\,\underline{x}\,(t) + B\,\underline{w}\,(t) \quad ;$$

(50): $$[x,y,z,b]^T = \Gamma\,\underline{x}(t)$$

For the state-vector definition of (49)-a, the coupling matrix, A, will be upper-triangular, and this will result in some computational advantages in terms of 11 by 11 matrices. For the state-vector definition of b), the generally 11 by 11 vector-matrix equation of (49)-b breaks apart into 4 independent equations, three of which are 3 by 3 and the one involving b is 2 by 2.

Under the standard Kalman assumptions, the data input to the filter, z(k), consists of coordinate to be tracked, x(k), plus noise, n(k). It is assumed that the noise, n(k), is independent of j(k), is zero-mean, white, Gaussian, and of known variance, $(\sigma_n)^2$. Then, the data form is

(51): $$z(k) = x(k) + n(k)$$

The Kalman filter equations are given below together with the generator equations and data form, for completeness

$$\underline{\text{Generator}}: \quad \underline{x}(k+1) = \underline{\Phi}\,\underline{x}(k) + \underline{\gamma} j(k)$$
$$x(k) = \underline{\lambda}^T \underline{x}(k) \qquad ; \underline{\lambda}^T = [1,0,0]$$

(52): $\underline{\text{Data}}: \quad z(k) = x(k) + n(k)$

$$\underline{\text{Filter}}: \quad \hat{x}(k) = \underline{\lambda}^T \hat{\underline{x}}(k)$$
$$\hat{\underline{x}}(k) = \underline{g}(k)\,[z(k) - \underline{\lambda}^T \underline{\Phi}\,\hat{\underline{x}}(k-1)] + \underline{\Phi}\,\hat{\underline{x}}(k-1)$$

In(52), the quantity, $\underline{g}(k)$, is a gain vector, defined by

(53): $$\underline{g}^T(k) = [g_x(k), g_v(k), g_a(k)]$$

$\underline{g}(k)$ can be time-varying. Its method of calculation is detailed below.

To obtain understanding of the operation of the filter, the vector-matrix equations of (52) are written in coupled scalar form as

$$\begin{aligned}
\hat{x}(k) &= \hat{x}(k-1) + T\,\hat{v}(k-1) + (T^2/2)\,\hat{a}(k-1) + g_x(k)\,e(k) \\
\hat{v}(k) &= \hat{v}(k-1) + T\,\hat{a}(k-1) + g_v(k)\,e(k) \\
\hat{a}(k) &= \hat{a}(k-1) + g_a(k)\,e(k)
\end{aligned} \qquad (54)$$

$$e(k) = z(k) - [\hat{x}(k-1) + T\,\hat{v}(k-1) + (T^2/2)\,\hat{a}(k-1)]$$

Figure 13 shows the filter structure.

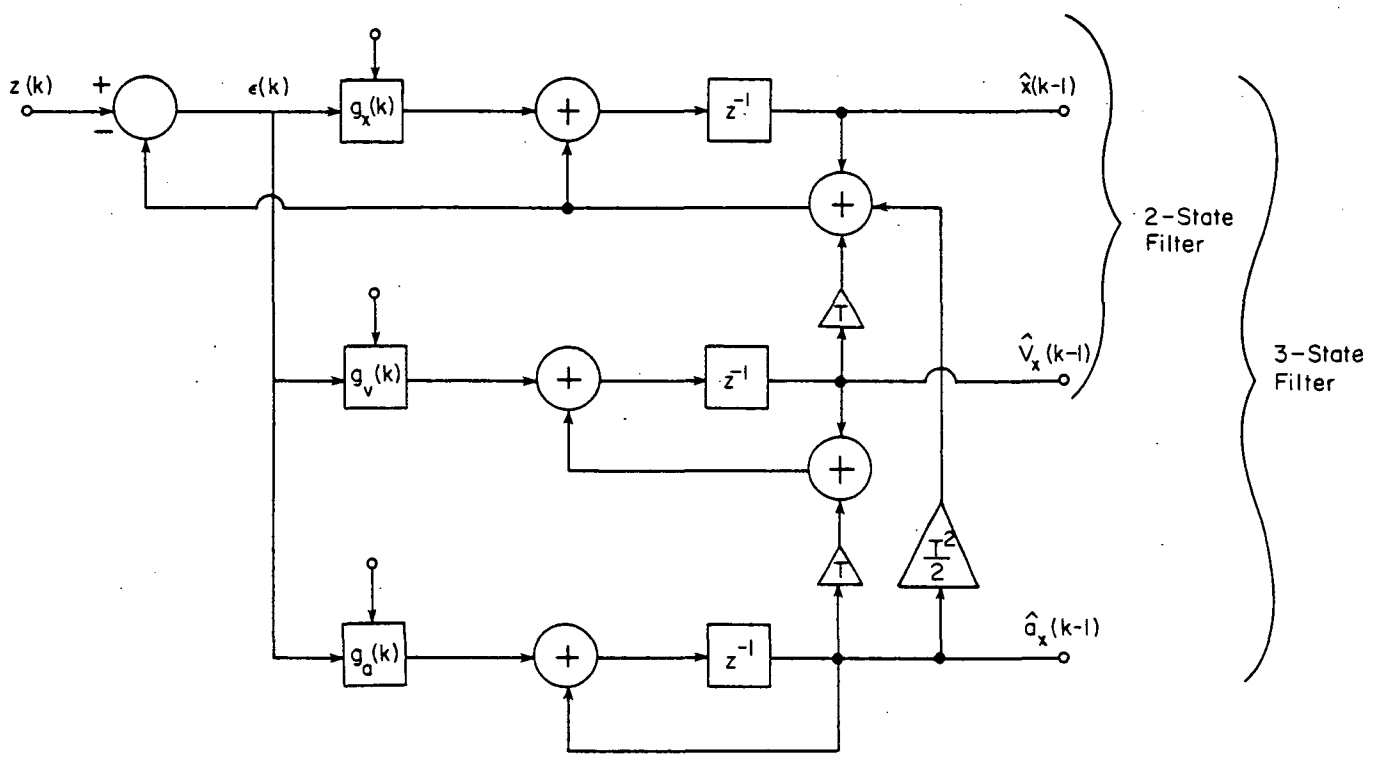

Figure 13. Kalman Filter.

It is seen from (54) and the figure that each new estimate consists of a prediction, based on the old estimates, plus a correction, based on the new data. The correction is a weighted error term where the error is data minus predicted position. The weighting factors are the "Kalman gains". The error term, e(k), is called "filter residual" or "innovation".

In the standard Kalman filter, the gain vector, $\underline{g}(k)$, is computed from a set of coupled vector-matrix equations given below. These equations are based on the covariance matrix of the filtering error, where

Filtering
Error : $\underline{\tilde{x}}(k) = \underline{x}(k) - \underline{\hat{x}}(k)$

Covariance
(55): Matrix : $v(k) = E[\underline{\tilde{x}}(k)\underline{\tilde{x}}^T(k)]$

$E[\ ]$: statistical average

under the assumption that $\underline{x}(k)$ and $\underline{\hat{x}}(k)$ both have zero mean values. The gain equations are

$$v(k|k-1) = \underline{\Phi}\, v(k-1)\, \underline{\Phi}^T + (\sigma_j)^2 \underline{\gamma}\,\underline{\gamma}^T$$

(56): $$\underline{g}(k) = v(k|k-1)\ \underline{\lambda}[(\sigma_n)^2 + \underline{\lambda}^T v(k|k-1)\underline{\lambda}]^{-1}$$

$$v(k) = [I - \underline{g}(k)\,\underline{\lambda}^T]v(k|k-1)$$

In (56), $v(k)$ is covariance matrix computed using the $\underline{\hat{x}}(k)$ filtered estimate, while $v(k|k-1)$ is the covariance matrix computed using the $\underline{\hat{x}}(k|k-1)$ one-step predicted estimate. $(\sigma_j)^2$ and $(\sigma_n)^2$ are the known variances of the modeled jerk and additive white noise processes, respectively.

The $\underline{g}(k)$ computation is recursive and is initialized with a value $v(k=0)$. Usually this is the steady-state variance matrix for the process, $\underline{x}(k)$. However, in the present navigation case, with the given structure of the $\underline{\Phi}$-matrix, the variance of $\underline{x}(k)$ grows without bound. Thus, some other "ad hoc" initialization should be used for $v(k)$.

All the gain elements of $\underline{g}(k)$ rapidly approach steady-state values. In the steady-state, the Kalman filter is just the optimum Wiener filter. The steady-state gains are just functions of T, the sampling period and the ratio $(\sigma_j)^2/(\sigma_n)^2$ = SNR which is a "signal to noise ratio". The greater the ratio, the greater the gains and vice-versa. The greater the gains, the more responsive the filter is to abrupt changes (maneuvering) in the input, but the more white measurement noise is accepted by the filter. The lower the gains, the more sluggish the filter becomes, but the less noisy becomes the output. Figures 14 and 15 [19] show the examples of the evolution of the gains and the

steady-state dependence on signal to noise ratio.

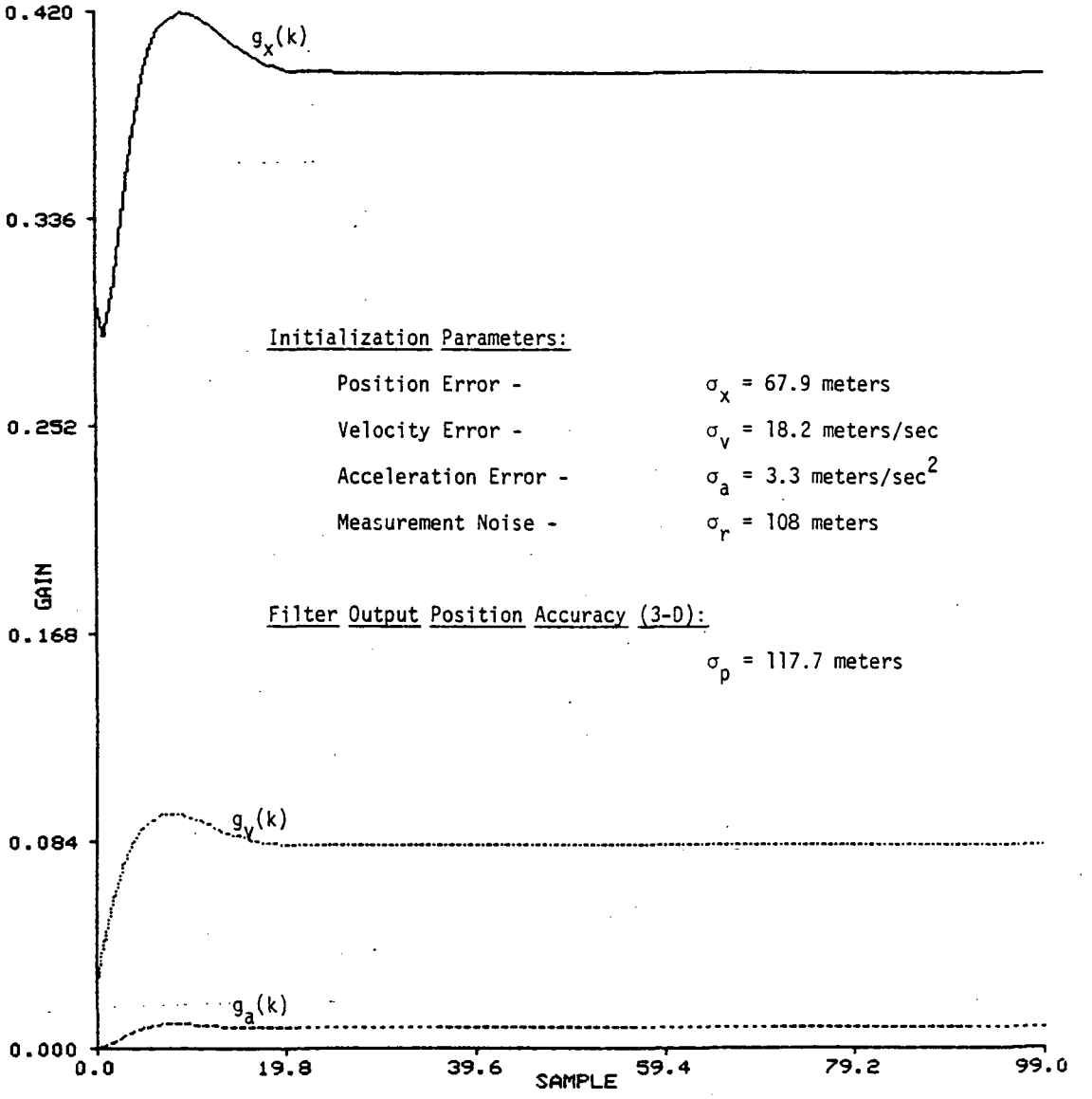

Figure 14. Evolution of Gains.

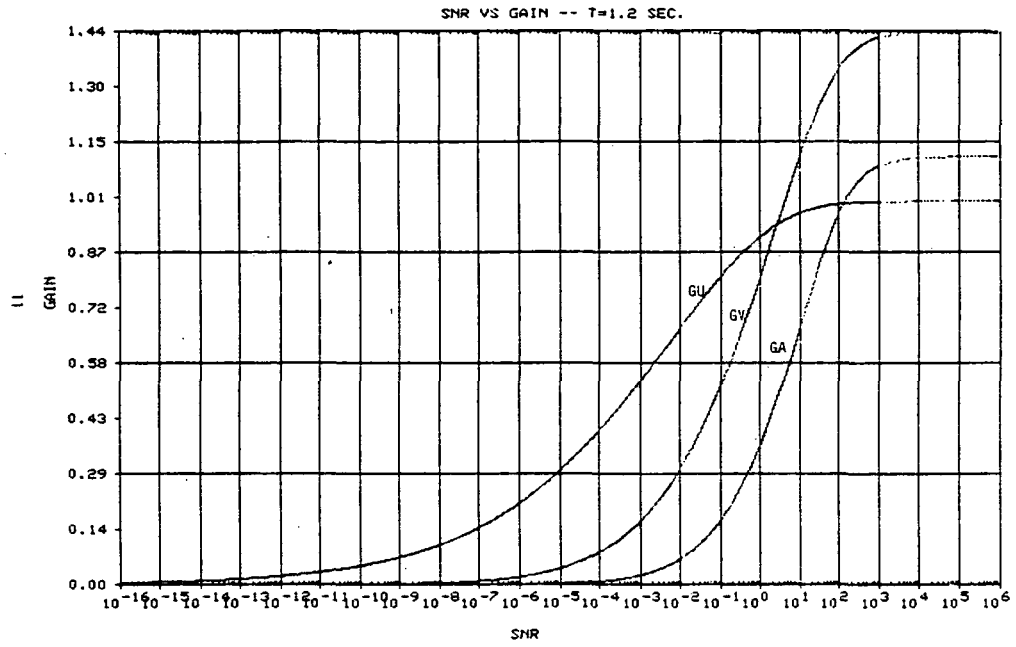

Figure 15. SNR Dependence of Gains.

A problem with using a Kalman filter in the navigation problem is that the SNR ratio changes drastically, dependent on vehicle maneuvering. During unaccelerated motion, SNR→0 and the optimum steady-state gains are very small, resulting in little noise on the filter output. During accelerated motion, SNR increases and the optimum steady-state gains are larger, resulting in more white noise acceptance but lower dynamic filter error.

For example, a 300 knot aircraft subject to + or - 3 meter normal control excursions with a 30 second period and 10.8 meter ranging noise, results in SNR = 6.5×10^{-6}. During a standard rate two-minute turn, SNR = 1.54×10^{-3}, a + 30dB change. The standard Kalman gain computation has no knowledge of these radically different environments for filter operation. Thus, in order to use the filter effectively, an "adaptive" gain algorithm must be employed.

There exists no well-developed body of theory for adaptive-gain filters, comparable to Kalman-Wiener theory. Thus, developements in this area are all "ad hoc" and for special cases. The only basic requirement is that the filter residual, e(k), be used to provide information for adjusting the filter gains.

The gains may not be adjusted independently. Rather, they must be adjusted consistent with results which are produced by solution of the gain equations. Otherwise, filter instability may result.

There is one final problem in applying the Kalman filter to the GPS navigation problem. That is the noisy navigation measurements, $R_i(k)$, do not constitute the correct data form for the Kalman filter. The filter requires noisy measurements of the position coordinates, themselves. The pseudo-range measurements, however, are non-linear transformations of the position coordinates. An approach to this problem is to use the linearization developed in section 6 and embodied in equation (41). That relation is expanded here as

$$(57): \quad \underline{x}(k) - \hat{\underline{x}}(k) = H^{-1}[\underline{R}(k) - \hat{\underline{R}}(k)]$$

In (57), $\underline{x}(k)$ is the true position vector, $\underline{x}^T(k) = [x,y,z,b]$, and $\hat{\underline{x}}(k)$ is the position assumed for purposes of computing H^{-1}. $\underline{R}(k)$ is the vector of four measured pseudo-ranges. $\hat{\underline{R}}(k)$ is the prediction of $\underline{R}(k)$, computed using $\hat{\underline{x}}(k)$ and the non-linear function of (44). Because the $\underline{R}(k)$ have additive white noise, so (57) will add white noise to $\underline{x}(k)$ through the linear transformation, H^{-1}. However, (57) gives a differential transformation from $\underline{R}$ to $\underline{x}$, rather than a global transformation. So the question remains, how to use (57) to provide the input data for the Kalman filters.

Suppose that in (57) at sample number k, the assumed position vector, $\hat{\underline{x}}(k)$, is made up of the Kalman filter predicted position components, which are all of the same form as that for the x-component, given as

$$(58): \quad \hat{x}(k|k-1) = \underline{\lambda}^T \underline{\Phi}\, \hat{\underline{x}}(k-1)$$

Now, we may write (56), showing explicitly the additive noise contribution due to $H^{-1}\ \underline{R}(k)$, as

$$\underline{x}(k) + \underline{n}(k) - \hat{\underline{x}}(k|k-1) = H^{-1}[\underline{R}(k) - \hat{\underline{R}}(k|k-1)]$$

$$(59): \quad = z(k) - \hat{\underline{x}}(k|k-1) = H^{-1}[\underline{R}(k) - \hat{\underline{R}}(k|k-1)]$$

$$= \underline{e}(k) = H^{-1}[\underline{R}(k) - \hat{\underline{R}}(k|k-1)]$$

That is, we may identify the quantity $H^{-1}[\underline{R}(k)-\underline{\hat{R}}(k|k-1)]$ as the 4-vector error term in an 11-state Kalman filter.

Because of our choice of the coupling matrix, $\underline{\Phi}$, in the generating model, that model reduced from an 11-state model to four independent models having 3,3,3, and 2 states respectively, based on the state vector description of (49-b). The Kalman filter may be similarly composed of four independent sections, each driven by one of the four elements of the error-vector, $\underline{e}(k)$. The extended Kalman Filter, so realized, is diagrammed in Figure [16].

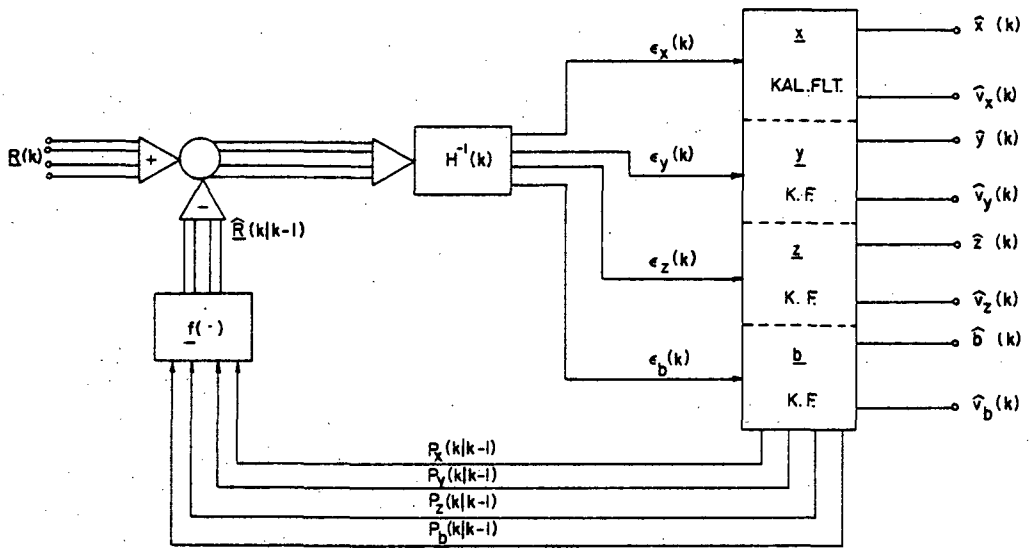

Figure 16. Extended Kalman Filter.

Although the structure of the Kalman filter is successfully de-coupled so far as computation is concerned, the position estimates are not independent, since errors in x, y, z and b are cross-coupled by f() and H^{-1}. Another effect is that the white noise variances in the inputs to the x, y, z, and b filters are not necessarily equal or fixed, since they are dependent on H^{-1}, which is variable with the geometry.

To illustrate the effect on the filter input noise, let H^{-1} be written as

$$(60): \quad H^{-1} = \begin{bmatrix} q_{x1} & q_{x2} & q_{x3} & q_{x4} \\ q_{y1} & q_{y2} & q_{y3} & q_{y4} \\ q_{z1} & q_{z2} & q_{z3} & q_{z4} \\ q_{b1} & q_{b2} & q_{b3} & q_{b4} \end{bmatrix}$$

Let the additive noise in $\underline{R}(k)$ be explicitly indicated by

$$(61): \quad \underline{R}(k) = \begin{bmatrix} R_1(k) \\ R_2(k) \\ R_3(k) \\ R_4(k) \end{bmatrix} + \begin{bmatrix} n_1(k) \\ n_2(k) \\ n_3(k) \\ n_4(k) \end{bmatrix}$$

Then, the noises effective in x, y, z, and b are given by

$$n_x(k) = \sum_{i=1}^{4} q_{xi}\, n_i(k)$$

$$(62): \quad n_b(k) = \sum_{i=1}^{4} q_{bi}\, n_i(k)$$

It is assumed that the range measurement noises, $n_i(k)$ are zero-mean, independent, and of equal variance, $(\sigma_r)^2$. Then, the variances of $n_x(k)$, $n_y(k)$, $n_z(k)$, and $n_b(k)$ are

$$(\sigma_{nx})^2 = (\sigma_r)^2 \sum_{i=1}^{4} (q_{xi})^2$$

$$(63): \quad (\sigma_{nb})^2 = (\sigma_r)^2 \sum_{i=1}^{4} (q_{bi})^2$$

The above delevopment of the extended Kalman filter has assumed that all four pseudo-range measurements are available at the same instant, and that the total filter is cycled with four available pseudo-ranges. This is, of course, not the case with a sequential receiver. Let us examine (57) in more detail

$$(64): \quad \begin{bmatrix} x-\hat{x} \\ y-\hat{y} \\ z-\hat{z} \\ b-\hat{b} \end{bmatrix} = \begin{bmatrix} q_{x1} & q_{x2} & q_{x3} & q_{x4} \\ q_{y1} & q_{y2} & q_{y3} & q_{y4} \\ q_{z1} & q_{z2} & q_{z3} & q_{z4} \\ q_{b1} & q_{b2} & q_{b3} & q_{b4} \end{bmatrix} \cdot \begin{bmatrix} R_1-\hat{R}_1 \\ R_2-\hat{R}_2 \\ R_3-\hat{R}_3 \\ R_4-\hat{R}_4 \end{bmatrix}$$

$$= \begin{bmatrix} q_{x1}(R_1-\hat{R}_1) \\ q_{y1}(R_1-\hat{R}_1) \\ q_{z1}(R_1-\hat{R}_1) \\ q_{b1}(R_1-\hat{R}_1) \end{bmatrix} + \cdots\cdots + \begin{bmatrix} q_{x4}(R_4-\hat{R}_4) \\ q_{y4}(R_4-\hat{R}_4) \\ q_{z4}(R_4-\hat{R}_4) \\ q_{b4}(R_4-\hat{R}_4) \end{bmatrix}$$

Equation (64) shows that the coefficient q_{x1} brings the differential range component $(R_1-\hat{R}_1)$ into the filter to correct the x-coordinate. Likewise q_{y1} uses $(R_1-\hat{R}_1)$ to correct the y-coordinate, etc. The first column of H^{-1} thus corrects all four position coordinates from the information supplied by the $(R_1-\hat{R}_1)$ differential range measurement. Likewise, the second, third, and fourth columns of H^{-1} apply the position vector corrections due to the second, third, and fourth satellites, repectivley.

An interesting question is, "What if these corrections are applied sequentially, rather than simultaneously?" That is, what if the filter is cycled every time a single new range measurement is obtained? The answer is that such a procedure works well, provided that the sampling rate is high compared to the maneuvering rate of the vehicle. As an example, Figure [17] shows the actual output of a sequentially-corrected Kalman operating with GPS (C/A only) on-board a ship.

MEAN Z-SET POSITION IS 38.42383 BY -78.65839 DEGREES

[[[[[Z-SET DATA ONLY]]]]]

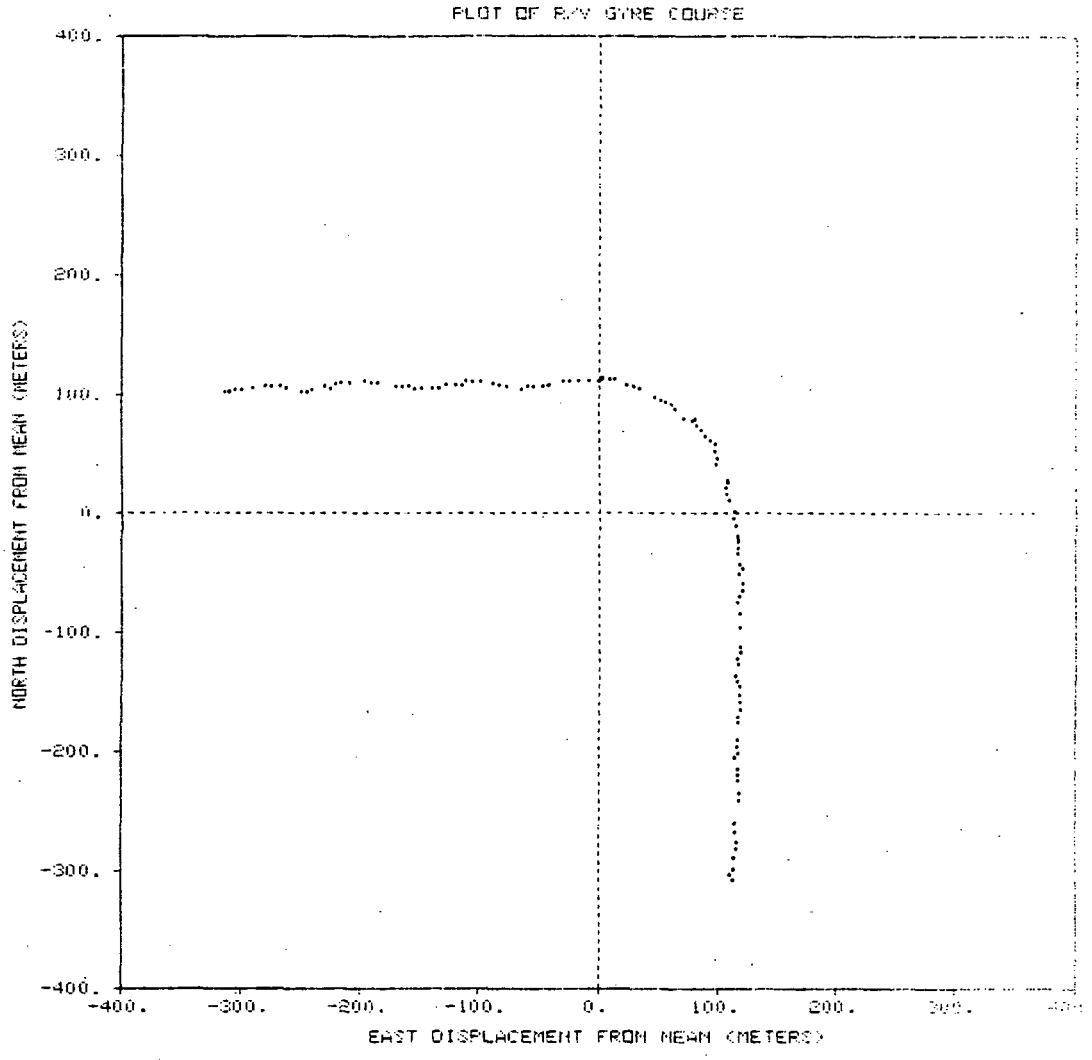

Figure 17. Sequential Receiver Track.

In the figure, the vessel initially steams east at ten knots (5.16 meters/sec). It then turns south smoothly, executing the 90-degree turn in 55 seconds. The range sample rate is one range every 1.2 seconds. The track displays little random noise, roughly 6 meters, 1-sigma, and no noticeable overshoot on the turn. The sequential correction of the filter has no noticeable effect for this case.

8. Conclusion.

At the time of this writing, not much has been published concerning the actual performance of GPS. What little has been released by the military or gotten into print about civil experiments, shows the performance to be extraordinary. Actually, the geodetic positioning accuracy of the system is so great that there is no standard against which to measure it. It seems likely that GPS itself will become the standard for measuring geodetic and navigation accuracies.

The article by Henderson and Strada [20] was the first widely available "official" release of GPS test data. Other sketchy data has appeared at various technical symposia. The author, for example, has been a co-investigator on a civil GPS maritime experiment, from which interesting data is released from time to time.

The first really informative performance result was that the C/A-only performance is not ten times worse than p-code performance, notwithstanding the factor of 10 difference in code resolution. Were the L2 frequency to be standardized for C/A code, so that a C/A-only user could have the ionospheric correction, it is likely that P-code and C/A-code performances would be commensurate for some users.

For surface navigators, the ionospheric perturbations are not particularly troublesome, so long as three satellites are available with good spread in azimuth and reasonably high elevation angles. For this kind of case, the horizontal ionospheric biases tend to cancel, leaving the main contribution in altitude. With 18 operational satellites, a surface navigator can construct a satellite selection algorithm to choose satellites as mentioned above, in order to minimize horizontal ionospheric effects.

Performance data from [20] shows P-code absolute navigation accuracies of 2.84 meters bias and 5.47 meters noise (1-sigma) in surface navigation at low velocities for one user set. Another P-code set in low-velocity surface navigation showed 2.84 meters bias and 6.4 meters noise (1-sigma) absolute. Helicopter flight tests yielded P-code performances of 9.6 meters bias and 4.7 meters noise (1-sigma) in three dimensions. All of these P-code results were

taken over short periods of time, less than 20 minutes, in duration.

One C/A-code-only result was given in [20], over a duration of 6 minutes. This showed bias of 30.3 meters and noise of 9.2 meters (1-sigma) for 3-dimensional navigation of a medium speed aircraft. Of the 30-meter bias, 29.0 meters was in altitude. The horizontal accuracy of that flight performance was 8.6 meters bias and 8.3 meters noise (1-sigma), absolute.

C/A-code-only results for a ship docked either at Galveston, Texas or Miami, Florida have consistently shown biases less than 30-meters horizontal and noises of 6-meters, (1-sigma) absolute, over 2-hour time intervals. The biases are based on the best small-scale harbor charts available. These latter C/A results were best-performance cases.

Although carefully calibrated performance results for GPS were scarce at the time of this writing, many GPS evaluations were underway at that time. it is likely that by the time of publication of this work, that many more results will be in the open literature. This author conjectures that new performance results will continue to show the extreme precision available from GPS, both with P-code and C/A-only.

REFERENCES

[1] Richard H. Battin, Astronautical Guidance, McGraw-Hill Book Co., 1964.

[2] P.R. Escobal, Methods of Orbit Determination, John Wiley & Sons, 1965.

[3] Thomas A. Stansell, The Transit Navigation Satellite System, Magnavox Report R-5933, Torrance, California, 1978.

[4] T.O. Sepplin, The Department of Defense World Geodetic System 1972, The Defense Mapping Agency, Washington, D.C., May, 1974.

[5] J.A. Klobuchar, Ionospheric Effects on Satellite Navigation and Air Traffic Control Systems, Air Force Geophysics Laboratory, Hanscom AFB, MA 01731, p. 3.

[6] K.W. Yip and O.H. Von Roos, "A New Global Ionospheric Model", JPL Deep Space Report 42-30. Jet Propulsion Laboratory, Pasadena, CA, p. 2.

[7] Kenneth Davies, Ionospheric Radio Propagation, Monograph 80, National Bureau of Standards, 1965, p. 6.

[8] Mitchell D. Eggers, Ionospheric Delay in the GPS, TCSL Tech. Report 81-06, Telecommunication and Control Systems Laboratory, Dept. of Electrical Engineering, Texas A&M University, May 19,1981, p. 44.

[9] J.H. Painter, "Designing Pseudo-Random Coded Ranging Systems", IEEE Trans. Aerospace and Electronic Systems, vol. AES-3, No. 1, Jan. 1967, pp. 14-27.

[10] C.A. Bartholomew, "Satellite Frequency Standards", Global Positioning System, The Institute of Navigation, Wash. D.C., 1980, pp. 21-28.

[11] Solomon W. Golomb, Shift Register Sequences, Holden-Day, Inc., 1967, p. 33.

[12] J.J. Spilker, "Signal Structure and Performance Characteristics", Global Positioning System, The Institute of Navigation, Wash. D.C., 1978, pp. 29-54.

[13] S.W, Golumb, L.D. Baumert, M.F. Easterling, J.J. Stiffler, and A.J. Viterbi,Digital Communications With Space Applications, 1964, pp. 85-105.

[14] H.W. Farris, M.P. Ristenbatt, et. al., An Introduction to Pseudo-Random Systems, vol. I - Basic Concepts and Techniques, Tech. Report 104-1, the Cooley Electronics Laboratory, University of Michigan, 1960, Classified CONFIDENTIAL (for some obscure reason).

[15] Phil Ward, "An Advanced NAVSTAR GPS Multiplex Receiver", Record of the IEEE 1980 Position, Location, and Navigation Symposium, Atlantic City, N.J., Dec. 8-11, 1980, pp. 51-58.

[16] W.C. Lindsey and M.K. Simon, Telecommunication Systems Engineering, Prentice-Hall, 1973, pp. 62-64.

[17] P.S. Noe and K. Bennett, Precision Analysis for Low-Cost GPS Receiver Micro-Processors, DSL Report 81-(03), Digital Systems Laboratory, Elec. Eng. Dept., Texas A&M University, May, 1981.

[18] Hamed Parsiani, Navigation Algorithm for Single Channel Low-Cost GPS Receiver, Ph.D. Dissertation, Dept. of Elec. Eng., Texas A&M University, August, 1979, pp. 28-45.

[19] Steven B. Hyde, In Regard to the GPS Adaptive Gain Dilemma, TCSL Memo 81-05, The Telecommunication and Control Systems Laboratory, Elec. Eng. Dept., Texas A&M University, April, 1981, pp. 11-22.