

Appendix A

Expected Positioning Accuracy of Global Positioning System on Student Rocket Project 4

Mark C. Charlton

For

EE656 – Professor Hawkins

December 2000

ABSTRACT

The purpose of this investigation is to assess the positional accuracy that may be obtained using a single-frequency Global Positioning Satellite (GPS) system launched aboard a sounding rocket. It is intended as supporting analysis for Student Rocket Project 4 (SRP-4) at the University of Alaska-Fairbanks (UAF). Specifically, it concentrates on high-latitude sounding rocket launches and the errors expected in this unique environment when using a C/A-code receiver and code-based techniques. It includes a brief introduction to GPS and a history of GPS on sounding rockets launched from Poker Flat Research Range, located at approximately 65 degrees north latitude. Next it examines the various error contributions in terms of source and magnitude for both stand-alone and differential systems. This examination is followed by a summary of areas that warrant further research. Conclusions are drawn regarding expected performance, and possible approaches to achieve better accuracy are given.

A.1 INTRODUCTION TO GPS ON SOUNDING ROCKETS

A.1.1 POSSIBLE BENEFITS OF USING GPS ON SOUNDING ROCKETS

This analysis is being undertaken to evaluate the expected accuracy of a GPS tracking system on the SRP-4 sounding rocket. Before diving into an evaluation of this specific application, the reasons for attempting GPS tracking on sounding rockets in general should be evaluated. The tracking of sounding rockets started out using visual methods and progressed to radar techniques as the rockets began to travel beyond visual range. While radar has proven effective in most situations, it has limitations in some respects and it is in these areas that the Global Positioning System (GPS) stands to provide substantial benefits.

First, GPS has the potential to drastically reduce the cost of tracking sounding rockets and their payloads. The radar units currently in use are very complex, very expensive, and maintenance intensive. In fact, as they age, these radar sets stand to become unsupportable. The logistics costs to bring the radar and the operators to a range location are high, especially for reasonably remote locations such as Poker Flat Research Range, near Fairbanks, Alaska. In addition to the hardware, software, maintenance, and logistics costs, the per-hour personnel costs for highly trained technicians to operate the radar is substantial. This is especially true when these personnel need to be kept at the range during periods of inactivity while awaiting conditions suitable for launch. The magnitude of these costs strains the budget of current programs and is likely beyond the reach of many would-be sounding rocket experimenters. A GPS tracking system, by comparison, can be operated by the experimenters themselves with minimal training and requires very little beyond the existing infrastructure to implement.

Second, GPS stands to provide a marked increase in position and velocity accuracy over what is currently possible using radar. Most sounding rockets are tracked using a single radar unit that makes use of distance and angle measurements to determine position and velocity. The plots in reference [1] indicate that the two radar units used in that experiment do not agree with each other any better than they do with GPS. As noted by Mr. Barton Bull of NASA-Wallops, the radar distance measurement is the most accurate, while the elevation and azimuth angle measurements suffer errors due to mechanical tolerances and encoder errors [1]. A more accurate radar measurement can be made if you place multiple radar units at wide angles to each other and only use the distance measurements to triangulate for position, but this is complicated and expensive for all the same reasons mentioned above. While GPS has its own share of error sources, many of which will be evaluated in this paper, the resulting accuracy should be superior to radar in most cases.

Not only is GPS less expensive and less complicated than radar, it also has the associated benefit of being far more transportable. The flight unit is typically smaller than a handheld camera, and the ground portion may be implemented on the same computer that is used for telemetry retrieval. This makes GPS far more suitable for operating from ranges where radar is not only expensive to provide, but impractical or impossible. Mr. Bull describes this aspect of GPS as a strong motivation for NASA's movement away from radar and toward GPS [2].

Finally, GPS provides a mechanism for investigators to time-tag events on-board the rocket with incredible accuracy. The signal from each satellite is encoded with a precise time-tag relating the time of transmission to GPS time and therefore to UTC. As experiments become more sophisticated, and more accurate times are required, access to this precise timing signal will prove to be another major benefit of GPS over radar.

This section has highlighted a few of the significant advantages of using GPS for the tracking of sounding rockets. Having established that there are compelling reasons to implement GPS, the next section describes the basics behind how the system works.

A.1.2 BASIC CONCEPT OF OPERATION FOR GPS

The Global Positioning System (GPS) uses radiowave transmissions from a constellation of 24 earth-orbiting satellites to enable a user anywhere on the globe, or above its surface, to precisely determine his position and velocity in three dimensions. The GPS satellites are placed in a constellation comprised of six evenly spaced orbital planes each having an inclination of 55 degrees. Based on an optimization of the constellation, the individual satellites are unevenly spaced within their orbital plane to maximize the time that at least four are available to a ground-based user, even when a number of satellites are out of service. The individual spacecraft orbits are maintained with a semi-major axis of approximately 26561.75 km and an eccentricity of less than 0.02. This produces a near circular orbit with a period of 12 mean sidereal hours [3]. Figure 1 is an illustration of the GPS constellation and the orbital spacing of the satellites.

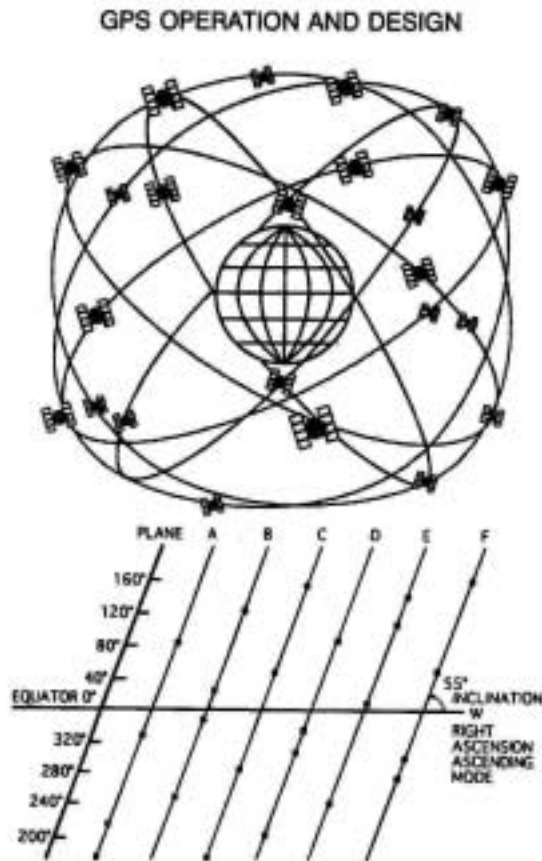


Fig. 1. Global Positioning Satellite and orbital spacing. From Fig. 6 of Reference [3].

The radiowave transmissions from these satellites are right-hand circularly polarized and are broadcast in the L-band at an L_1 carrier frequency of 1575.42 MHz and at an L_2 carrier frequency of 1227.60 MHz. They are composed of a spread-spectrum signal carrying a pseudorandom noise code. The detailed construction of the signal will not be discussed here, although references [3], [4], [5], and others examine this subject in great depth.

The fundamental principle behind GPS positioning is very straightforward. By measuring the angle or distance to three points of known location, a user can “triangulate” his position relative to those three points. This triangulated position may then be related to other coordinates, like those on a map, and expressed in convenient units such as degrees latitude and longitude.

A two-dimensional example of this process uses a compass and a map for land navigation. In this case the technique involves measuring a line-of-sight angle, relative to magnetic north, to each of two or more physical features that are also represented on a map. A line corresponding to each feature can then be drawn on the map at the measured compass heading. The intersection of the lines is the user’s position.

For the three-dimensional GPS process, the references are precisely located GPS satellites, instead of physical features. The user determines how far he is from each reference satellite. To get the distance to a satellite, he measures a time of travel for the signal transmitted by that satellite. The transmission time is “coded” into the signal itself and is extracted and compared to the time of receipt as measured by the user’s receiver clock. This measured transit time is multiplied by the speed of the signal to get a distance to the transmission source, the satellite. This is simply an application of:

$$d = v \times t \tag{1}$$

where: d = distance traveled,
 v = velocity of signal,
 t = elapsed time.

Unlike the line generated in the two-dimensional land navigation example, this creates a sphere centered on that satellite with a radius equal to the distance calculated. The user now knows he is somewhere on the surface of that sphere. Repeating the process using a second satellite produces another sphere, the intersection of the two spheres being a circle. A third satellite measurement produces a third sphere and the intersection of the three spheres is a set of two points on a circle, one of which is the user’s location. The correct point can usually be selected by having a general idea of the user’s location (i.e. one is obviously not correct and can be discarded). Figure 2 illustrates this positioning process. As in the land navigation example, if the

user knows the location of each reference satellite relative to some “map”, he can now transform his solved-for position into a position on that same “map”. In the case of GPS, every satellite transmits the precise location of itself and every other satellite as part of its signal. With this information, the user receiver is able to relate its position solution to the surface of the globe and output the position in terms of degrees latitude and longitude. This perfect position solution is, however, subject to a number of errors in its real-world implementation.

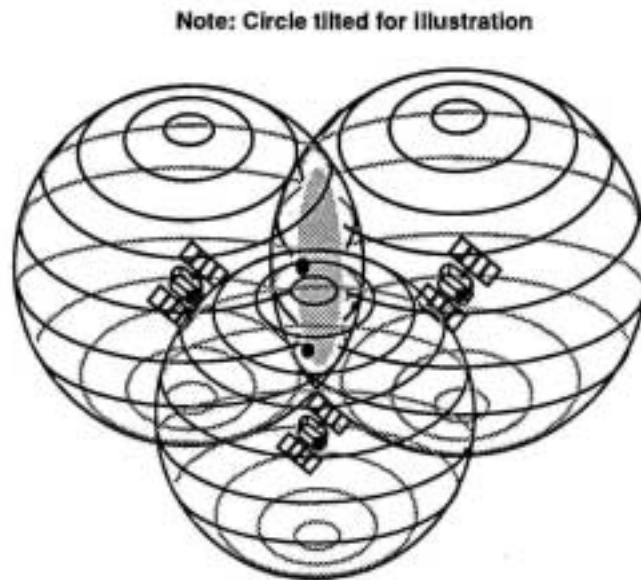


Fig. 2. Basic illustration of GPS positioning technique. From Fig. 2.6 of Reference [4].

A.1.3 WHY ARE THERE ERRORS IN A GPS POSITION SOLUTION?

While the GPS concept is relatively straightforward, this solution process relies on a number of assumptions that do not strictly hold true in the real-world implementation. First, the receiver must be able to precisely measure the transit time for each signal used in its solution. Second, the signals must travel at a known, constant speed. Finally, the signals must travel in straight lines from the satellites to the user receiver. Violations of these assumptions correspond to errors in distance measurement to the reference satellites. Distance errors contribute to inaccuracy in the final position solution, and an inspection of the simplified two-dimensional view in Figure 3 reveals one way of depicting their effect.

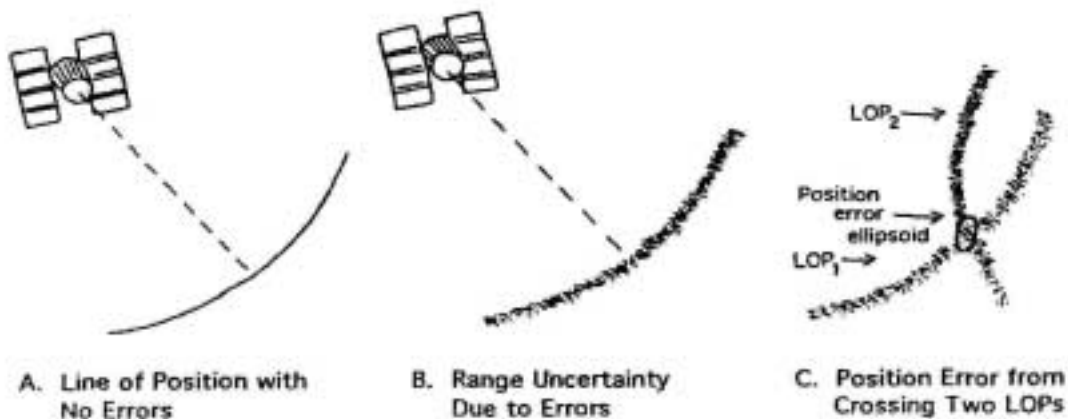


Fig. 3. Effect of error on lines of position. From reference [6].

It is seen that the intersecting lines are no longer “fine” lines, but lines plus or minus some uncertainty. The “thickness” of the lines of position is due to the aforementioned errors, which prevent the distances from being measured exactly. Extending this to the three-dimensional case, the thickness of the lines becomes thickness of the sphere surfaces. The end result is that the user only knows his position to be within the intersection of the spheres, and this intersection is no longer a point due to the thickness of the surfaces. When measuring the distance to each satellite, a larger error yields a “thicker” surface and a larger intersection of the surfaces. This, in turn, yields a greater uncertainty in the user’s position determination. Since these measurements are not the true distance to the satellites, they are commonly called *pseudorange*s, or *PRs*, in the literature and this term will be used throughout the remainder of this paper.

The first stated assumption indicates that the user must be able to accurately time the arrival of the signal in order to get a precise transit time measurement. The GPS spacecraft use very precise atomic time standards to time-tag the signal at transmission, but in order to keep the cost of user receivers manageable, the receivers use far less accurate crystal oscillators. These user receiver clocks have an offset from the true GPS time kept by the atomic standards and this bias would result in large position errors if uncorrected. To eliminate the bias, the actual implementation of GPS requires four satellites for a three-dimensional position solution. The solution algorithm becomes a problem of four equations in four unknowns, with the unknowns consisting of the three position dimensions and the clock bias. The measurement to a fourth satellite allows the determination of all four. This eliminates the largest error in measuring signal transit time. The small error due to inaccuracy of the atomic time standards and the errors in distance measurement due to violations of the remaining two assumptions will be discussed in later sections of this paper.

Position uncertainty is not only a function of distance measurement errors, but also of geometry. Referring again to Figure 3, it can be seen in the right-most diagram that lines of a given thickness crossing perpendicularly will yield a smaller area of intersection than those crossing at “shallow” angles. Recall that a smaller intersection corresponds to less uncertainty,

and is therefore desirable. The same is true in three dimensions, where a smaller volume of intersection corresponds to less uncertainty in the solution. Based on the satellites available, one combination will yield the best geometry to minimize the volume and therefore will yield the smallest position error. A measure of the "quality" of the geometry associated with a certain selection of satellites is known as Dilution of Precision, or DOP. A lower DOP corresponds to a smaller volume of intersection for the three spheres and therefore a smaller error or uncertainty.

A GPS user may be interested in solving for precise time, vertical position, horizontal position, or as in the SRP-4 case, three-dimensional position. For reasons that are beyond the scope of this paper, there are different varieties of DOP that correspond to the different desired solutions. Each has a different value and their computation requires a fairly complex mathematical process that is described in detail by Parkinson and Spilker [3] as well as others. For three-dimensional position, the DOP of interest is called the Position Dilution of Precision, or PDOP, and has a worldwide median value of approximately 2.5 [3]. The specific PDOP for a particular time and location may be calculated using the above-mentioned process, or can be easily obtained from an off-the-shelf software package as described by Kaplan [4]. Figure 4 is one example of a product from such a software package. This plot is generated for a particular location, at a specific time, and shows all the satellites in view above a pre-selected elevation angle. The program determines the four satellites that yield the best solution geometry, and these are indicated with squares. The resulting PDOP is shown in the lower left corner.

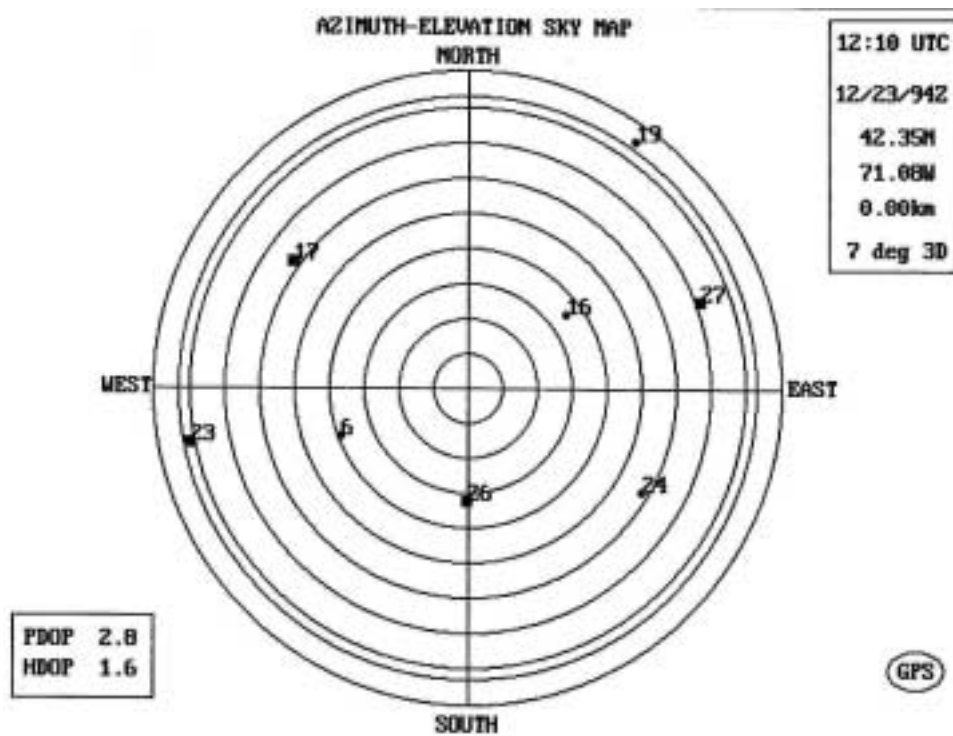


Fig. 4. Sample output for off-the-shelf GPS prediction software. From Fig. 7.30 of reference [4].

The same software can produce other useful products. One example is a plot that indicates the predicted satellite visibility over some period in the future, which might be useful for determining the worst-case expected coverage during a planned period of operation. Figure 5 is an example of such a plot. This product indicates which satellites are in view above a selected elevation angle at a particular location over a period of 24 hours. It also indicates if there will be any periods during which the PDOP exceeds a selected level, yielding an unsatisfactory level of error. A final example of a useful software product may be seen in Figure 6 which illustrates the PDOP as a function of time over a 24 hour period. This is generated for a particular location with a given mask angle.

The ability to predict the accuracy of a GPS-derived position is tied directly to an understanding of both the pseudorange errors and the specific PDOP for the satellites used in the solution. Given these two pieces of information, the user error can be estimated as [3]:

$$\text{position error} = \text{pseudorange error} \times \text{position dilution of precision.}$$

This basic calculation will be examined in greater detail in a later section. One objective of this investigation is to quantify this error for SRP-4 given the high-latitude environment in which the system will operate. When discussing GPS error, there are a number of typical measures of accuracy.

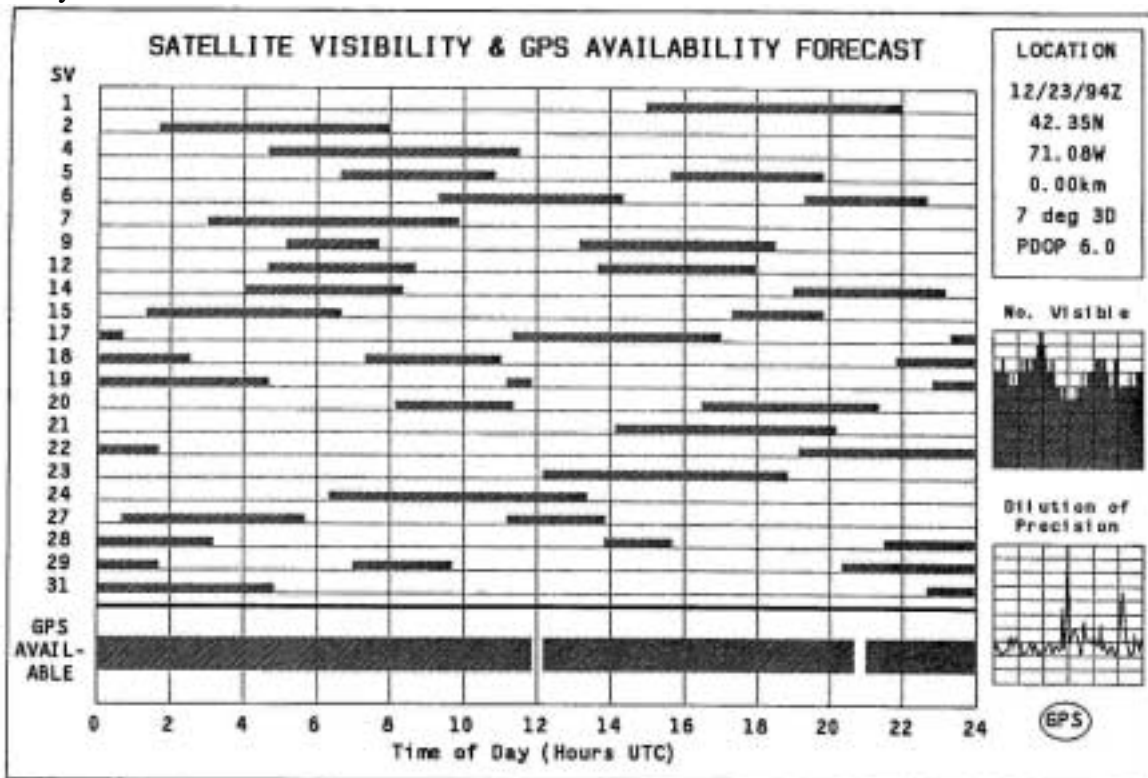


Fig. 5. Predicted satellite visibility/availability. From Fig. 7.32 of reference [4].

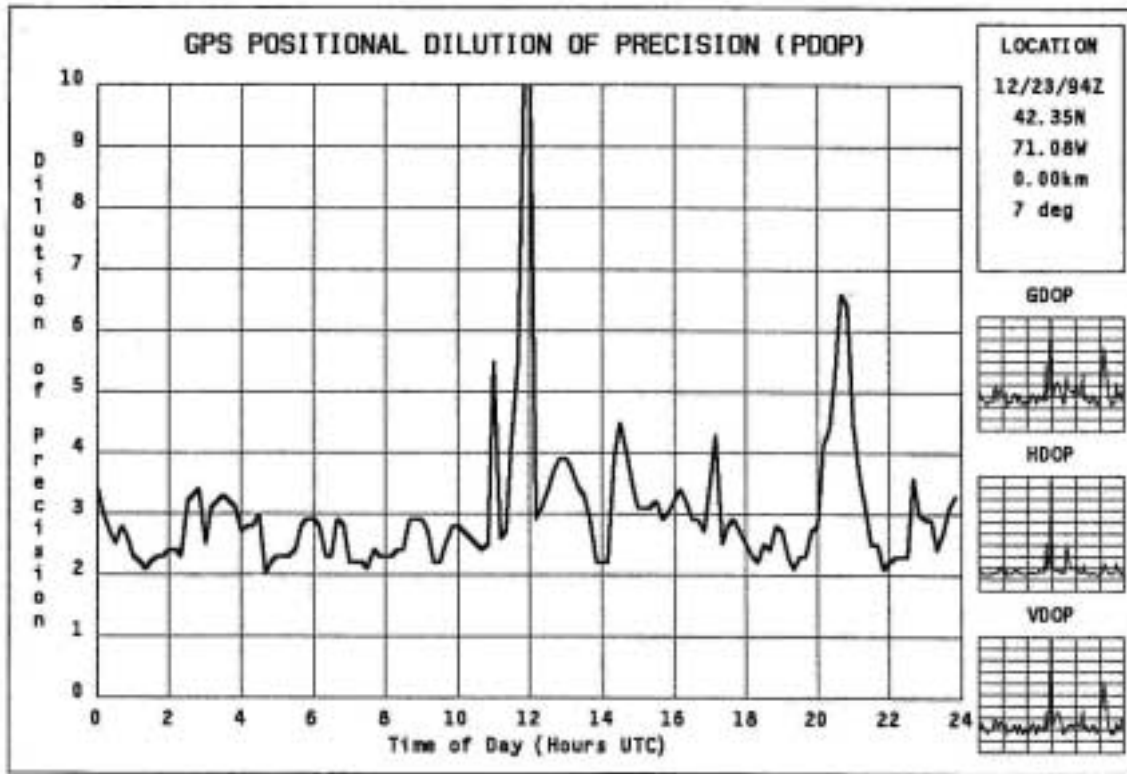


Fig. 6. PDOP profile over a 24-hour period. From Fig. 7.35 of reference [4].

A.1.4 TYPICAL MEASURES OF ACCURACY

Research into GPS performance turns up a number of "accuracy descriptors," many of which are not readily understandable. In his article, Frank van Diggelen [7] gives an in-depth explanation of the most commonly encountered measures and how to convert from one to another for comparison. As van Diggelen points out, the reason that so many different accuracy measures exist is that the performance of GPS varies statistically and therefore must be described by statistical means. To further complicate matters, there are one-, two-, three- and even four-dimensional measures of accuracy depending on what application of GPS a user is interested in. The statistical nature of the overall error in position is derived from the fact that many of the individual error contributions are time varying. These individual error components could be described in terms of their max or min values, but are typically treated as statistical distributions in order to get performance for a percentage of time.

For three-dimensional position errors, the most commonly used statistical measures are spherical error probable (SEP), and 1σ or root-mean-square (rms) error. The SEP gives the radius of a sphere, centered on the true antenna position, containing 50 percent of the points in

the three-dimensional scatter plot, or 50 percent probability. The rms position error is based on a root-sum-square (rss) of the 1σ pseudorange error contributions as will be explained in a later section. The terms “ 1σ ” and “rms” are often used interchangeably and taken to be equivalent. As pointed out by van Diggelen, this is only strictly true when the mean error is zero. Fortunately, the analysis of GPS error is simplified somewhat due to the fact that “the mean errors (over a sufficiently long time interval) are zero, or close to zero [7].” Therefore, in terms of the analysis in this paper, these two terms can be taken to mean the same thing. Further, based on the assumption that the errors are spherically scattered, the 1σ or rms error corresponds to a probability of 61-68 percent. The 2σ , also called the 2drms , error corresponds to a probability of 95 percent [7], [8]. While this assumption regarding the distribution is not strictly true, treating the errors in this manner “has proved to be surprisingly valid [3].” The rms, or 1σ , PR error is commonly referred to as the *user-equivalent-range-error*, or *UERE*. The predicted UERE for the SRP-4 system will be calculated in a later section of this paper.

Having established the measurements of accuracy typically used, it is important to point out that the magnitude of the predicted error for a system can vary substantially depending on the manner in which GPS is employed. The two broad categories of systems are often called stand-alone and differential.

A.1.5 STAND-ALONE VS. DIFFERENTIAL GPS

Stand-alone GPS refers to a receiver that is measuring pseudoranges to the satellites in view, determining those with the best geometry, and calculating a position and velocity solution without any other input from outside sources. A common example of this type of system is the hand-held receiver sold in many sporting goods stores.

Differential GPS (DGPS) is a more involved, and theoretically more accurate, implementation. A DGPS relies on corrections from some outside source to reduce error and improve the accuracy of the solution. Typically, this type of system uses a second receiver placed at a precisely surveyed location. This reference receiver compares its GPS-derived position to the known position provided by the survey. It then calculates either an overall position correction vector, or a correction for the pseudorange to each satellite in view. Most often, the method that corrects each pseudorange yields a more accurate solution. The overall position correction vector, or the individual correction for each pseudorange, is then applied to the mobile receiver’s measurements to get a differentially corrected solution. In the most common DGPS implementations, the corrections are broadcast to the mobile receiver and the solution is corrected in that receiver. The SRP-4 DGPS is implemented in a slightly different fashion. In the case of SRP-4, the pseudoranges measured by the mobile receiver are transmitted to the ground station. The corrections are calculated by the reference receiver, and then applied to the mobile receiver’s measurements at the ground station, instead of on-board the rocket. This has come to be called “inverse” DGPS [1], and this terminology will be used throughout this paper.

Both stand-alone and DGPS produce remarkable position accuracy, but neither will give a

perfect solution. The stand-alone units are subject to a number of errors and their overall position accuracy is attributable directly to the magnitude of these contributions. A differential system is able to correct errors that are correlated with time and distance, but some GPS errors are decorrelated to varying degrees and cannot be totally corrected. Despite this limitation, DGPS should be able to generate a better solution than a stand-alone unit. The accuracy of the final solution is a function of how large the uncorrelated errors are, how old the corrections are, and how far away the mobile receiver is from the reference receiver. It is also important that the DGPS reference receiver be able to see the same satellites as the mobile receiver to generate an optimum correction. The expected errors for each type of system will be discussed in greater detail later in this paper.

A.2 HISTORY OF GPS ON SOUNDING ROCKETS LAUNCHED FROM POKER FLAT RESEARCH RANGE

A.2.1 PAST MISSIONS

As a starting point for assessing the expected accuracy of the SRP-4 system, a review of past sounding rocket launches from Poker Flat Research Range was conducted. As of this writing, there have been several sounding rockets launched from Poker Flat with a GPS receiver on board. The University of Alaska-Fairbanks Student Rocket Project launched a GPS receiver on board the SRP-2 rocket in May 1995. More recently, the NASA-Wallops group has launched several sounding rockets from Poker Flat with GPS receivers on-board. At least one of these experiments was intended to measure distance between parts of the payload as they separated from one another, as opposed to determining an absolute position relative to a “map”. The NASA system used on the other flights is designed to determine absolute position and is very similar in concept to that developed for SRP-4. It is, however, much more expensive.

A.2.2 RESULTS TO DATE

The similar missions to date have not produced any definitive accuracy measurements. Unfortunately the sole UAF attempt, SRP-2, had an in-flight problem and the electrical payload unexpectedly shut down shortly after launch. It failed to return any useful data. The NASA system has been extensively tested to insure proper operation, and has been launched from ranges all over the world, including from Poker Flat [1], [2]. NASA has certified the system for operational use, but despite routinely returning “real time tracking data for between 95 and 100% of each mission’s flight ... two flights experienced undetermined failures resulting in the loss of all or a significant portion of the data [1].” Unfortunately, one of these test flights that lost data was a flight from Poker Flat that might have provided data to aid in this analysis. While its users currently estimate that the system provides better than 10 m position accuracy, they have been unable to make a conclusive measurement of the system accuracy. This lack of a definite

measurement, despite the many test and operational launches, highlights the difficulty of verifying the accuracy of a GPS solution in a sounding rocket application. In order to measure the accuracy, you must have something more accurate against which to compare it. In the past, radar measurements have provided this standard, however, GPS is expected to be more accurate than the radar solution that uses both range and angle measurements from a single radar. The NASA experiments seem to indicate that this is indeed the case. Quantifying how much better has proven difficult. As described in the first section, it may be possible to generate a more accurate radar solution by “tracking with three or more widely separated radars and using only the range measurements [1].” The GPS solution could then be compared to this more accurate GPS solution to assess relative accuracy. Plans to conduct such an experiment are currently being evaluated by NASA.

Due to the notable lack of accuracy data from a GPS system launched out of Poker Flat that is similar to SRP-4’s, the assessment of predicted performance for the SRP-4 flight GPS is based on an error budget derived from a theoretical analysis. The following sections of this paper describe the operational environment, the error sources, and expected error budgets for the SRP-4 system.

A.3 PREDICTED PERFORMANCE OF STAND-ALONE C/A-CODE GPS

The following analysis investigates the predicted PR errors for the flight GPS system when operated as a stand-alone unit, without differential corrections. The prediction of error for a particular flight is difficult due to the time variability of the conditions under which the system might be operated. Due to this time-varying nature, the position accuracy of the system is best described using statistical measures. Each of the significant error sources is evaluated with respect to the cause of the error, the expected range of values, and the expected mean value. It should be noted that this analysis does not include the effects of Selective Availability (SA). Selective Availability is a method by which the navigation signal is intentionally degraded to deny the full attainable accuracy to “unauthorized” users, often resulting in 1σ position errors of 50 m or more [3], [4]. This was, by far, the most significant source of error for single-frequency stand-alone GPS users, but it has recently been permanently discontinued in accordance with U.S. government policy. A summary of the remaining significant error contributions is included in Table 1 at the end of this section and is followed by an assessment of the predicted position accuracy for the stand-alone system.

A.3.1 PREDICTED PSEUDO-RANGE ERRORS

A.3.1.1 SATELLITE CLOCK ERROR

As mentioned earlier, the signal is time-tagged on-board the satellite using a very precise

atomic time standard. Despite the incredible accuracy of these clocks, even the very small residual error yields a measurable distance error when multiplied by the signal velocity of 3.0×10^8 m/s. The cesium and rubidium atomic standards used by GPS have errors on the order of 1 part in 10^{13} over a day [3]. This corresponds to a time error of 8.64×10^{-9} seconds and, from equation (1), a corresponding distance error of 2.6 m due to the clock error. To minimize this error, a clock correction factor is normally uploaded to each satellite once a day by controllers and this correction is part of the message broadcast by the satellites. As a result, the clock error is typically at a minimum immediately following this upload when the correction is most accurate. Based on a 1984 study referenced by Parkinson and Spilker [3], the observed error grew to approximately 4.1 meters over the 24-hour period between uploads. Kaplan [4] reports the 3.0 m value from the 1991 GPS Joint Program Office document, but does not indicate what point in the 24-hour upload cycle this value corresponds to. Since the standard deviations of the measurements have been observed to grow quadratically over time, a mean value of 1-2 meters is expected [3], and a typical value of 2.1 m is used in the table.

A.3.1.2 EPHEMERIS ERROR

As part of the positioning process, the receiver must have accurate references in relation to which it can derive its intersecting lines of position. The receiver measures only transit times for the signals and relies on accurate knowledge of the source satellites' positions to triangulate its own position. Positions for the reference satellites are determined by the control segment, uploaded to the satellites, and broadcast to users as part of the navigation message. An ephemeris error occurs when one or more of these broadcast locations is in error. Since the orbit of each satellite changes as time goes on, this error tends to grow with increasing time since the last control segment upload. Based on the same study referenced above, the rms ranging error attributable to ephemeris was 2.1 m for predictions up to 24 hours. Kaplan [4] references a 1991 GPS Joint Program Office document that reports the 1σ value to be 4.2 m. This more recent value will be used in the table of predicted PR errors.

A.3.1.3 IONOSPHERIC ERROR

With the demise of SA, ionospheric effects are typically the largest remaining source of error. Since this is the largest effect, and the one most likely to be different for high-latitude operations, it will be developed in much more detail than the other errors discussed in this paper. To understand this significant impact on the accuracy of a pseudorange measurement, an understanding of what the ionosphere is, how it is characterized, and how it effects radio signals is helpful.

General description of the ionosphere

As described by Allnut, "The ionosphere is a region of ionized plasma that extends from roughly

50 km to 2000 km above the surface of the Earth [9].” Many of the atoms in this region are ionized, primarily by electromagnetic radiation from the sun and to a lesser extent by cosmic radiation coming from deep space. Concurrent with this ionization process, is a continuous recombination process, and the two result in a changing net level of ionization. The ionized atoms and electrons are found in layers or regions characterized by their height above the surface and referred to by a letter designation. Figure 7 is one way of representing the typical heights and electron density values for the various layers.

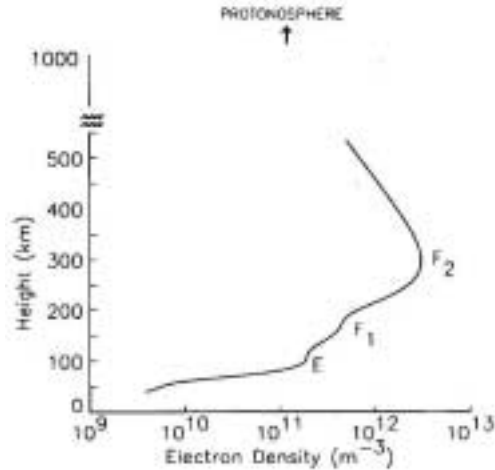


Fig. 7. Electron density of the different regions of the ionosphere vs. height for daytime conditions. From Fig. 1 of reference [3].

Figure 8 gives another graphical representation in which it is clear that the daytime and nighttime ionosphere are different, with some regions disappearing at night.

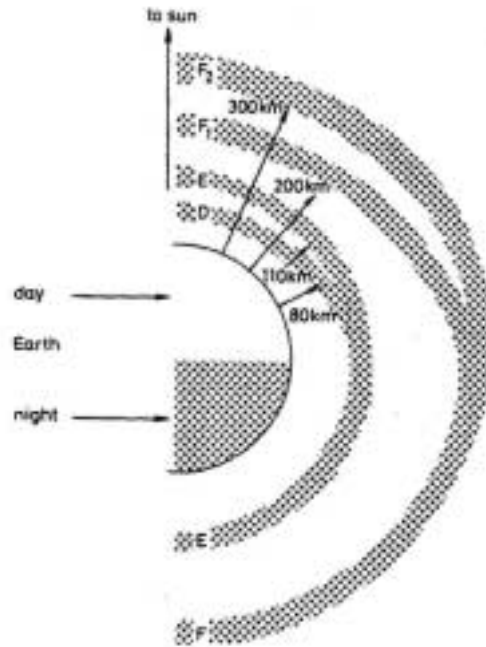


Fig. 8. Regions of the ionosphere as a function of height. From Fig. 2.5 of reference [9].

This may be expected since the primary mechanism for ionization is solar radiation. At night, with no incident radiation, the recombination process outpaces the ionization process and the result is a lower net ionization. Indeed, as seen in the diagram, the D region and one of the F regions more or less disappear. As pointed out by Allnut [9], however, during periods of intense solar activity, remnants of these regions may remain during the night. In addition to being dependent on diurnal cycles, the level of ionization is a function of the 11-year solar sunspot cycle. Increased sunspot number corresponds to higher solar radiation levels, and therefore to higher ionization and a more active ionosphere. Intense solar storms are yet another mechanism that may result in a more active ionosphere and higher levels of ionization. This may be one mechanism that causes the high-latitude ionosphere to vary significantly from the "typical" values used in this analysis.

Ionization yields free electrons and the presence of these electrons in the ionosphere makes it a "dispersive" region. This means that the effects on a radiowave transiting the ionosphere are a function of frequency as well as a function of the number of free electrons present. Since the transmitted frequency is known, only the number of electrons encountered along the path must be determined to predict the effects on the signal. In practice, electron density is difficult to measure, so most discussions of ionospheric effects use Total Electron Count (TEC) as a parameter in place of the electron density, N . This parameter is more easily measured than electron density and is defined as the integrated number of electrons on the path. It is typically given as a zenith value in electrons/ m^2 for a path cross-section of $1m^2$ [4]. If the path is not vertical, the TEC must be scaled to account for the actual path angle relative to zenith. Figure 9 shows the geometry for this situation when the ionosphere is treated as a thin layer with

a mean height. It defines z_o as the angle measured from zenith to the satellite at the user's position. The angle z' is measured from zenith to the satellite at the ionospheric "insertion point," IP. The IP is the point where the path passes through the mean value for the height of the ionosphere, h_m . Finally R_E is the mean radius of the earth.

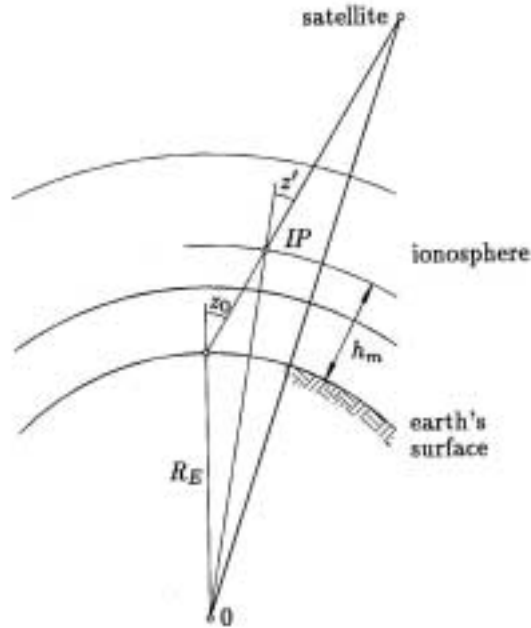


Fig. 9. Geometry for determining path Total Electron Content. From Fig. 6.2 of reference [5].

Using these definitions, the angle z' may be calculated using [5]:

$$\sin z' = \frac{R_e}{R_e + h_m} \times \sin z_o \quad (2)$$

where:

- z' = angle measured from zenith to satellite at insertion point,
- R_e = mean radius of the earth,
- h_m = mean value for the height of the ionosphere,
- z_o = angle measured from zenith to satellite at user location.

And the scaling equation to find TEC along a non-vertical path is [5]:

$$TEC_{path} = \frac{1}{\cos z'} \times TEC_{zenith} \quad (3)$$

where: TEC_{path} = Total Electron Count along the actual path in electrons/m²,

z' = angle from zenith,
 TEC_{zenith} = Total Electron Count on a vertical path in electrons/m².

Studies to quantify “typical” values of TEC are ongoing. Since TEC is directly tied to ionization, and ionization varies, we’d expect TEC to vary. Indeed, the “TEC is a function of time of day, user location, satellite elevation angle, season, ionizing flux, magnetic activity, sunspot cycle and scintillation. It nominally ranges between 10¹⁶ and 10¹⁹ electrons/m² with the two extremes occurring around midnight and mid-afternoon, respectively [6].” These times of maximum and minimum are to be expected since solar radiation is the primary ionizing mechanism. “No value of TEC greater than 10¹⁹ electrons/m² has ever been measured, even along a slant path in the Earth's ionosphere during a high solar maximum period, so that number is probably a safe maximum design value [3].” In addition to the small-scale fluctuations due to the mechanisms described above, there are localized irregularities that can cause large, rapid, and unpredictable changes in electron density and therefore TEC [9].

The fact that the level of ionization, and therefore the TEC, is not constant makes predicting effects on a particular signal difficult. The “typical” values can be used to bound the problem, but a more accurate knowledge of the TEC along the propagation path, at the time of transmission, is necessary to accurately predict the effects for a particular application. These more accurate values may be obtained by measuring the TEC in real-time, or by using a model.

Since ionospheric effects are frequency dependent, if the message is transmitted at two frequencies, a user should be able to measure one of the ionospheric effects and back out a TEC value. This “measured” TEC could then be used to remove any ionospheric effects. The GPS message is transmitted on both L₁ and L₂, and the process by which a dual-frequency GPS receiver can accomplish this correction is discussed near the end of this section. It should be noted, however, that most GPS users only have access to the signal on one of the L-band frequencies, and therefore use a single-frequency receiver and cannot make “dual-frequency” corrections. This is certainly true for the SRP-4 receiver. Lacking this ability to use the GPS signal itself to calculate the TEC, a measurement independent of GPS could be made and the results used to correct the position solution. A number of methods to measure TEC along a path have been demonstrated, but are generally impractical for the real-time correction of GPS position solutions. The Global Positioning System uses 24 orbiting satellites and a nearly infinite number of user locations, with each user-satellite combination having its own TEC_{path} value. For the worldwide problem, the sheer number of geometric variations makes measuring each impractical. For the more limited SRP-4 area of operation, there may be some feasible measurement method that would provide near-real time estimates of electron content along the path to the satellites used for its position solution, but this would be a complex endeavor.

Despite the difficulty of measuring a TEC value for every path, a model may be employed for day-to-day operations in an attempt to get a more accurate estimate for TEC than that given by the typical values. A number of ionospheric models exist and they are continually refined. A later section will examine the model currently used by single frequency GPS receivers to help estimate, and indeed correct for, some of the detrimental ionospheric effects analyzed in the next

section.

SPECIFIC EFFECTS OF THE IONOSPHERE ON THE GPS SIGNAL

Based on the foregoing discussion of how radiowaves are used for GPS positioning, the overview of the ionosphere, and the definition of the parameters used to characterize this region, this section evaluates the most significant ionospheric effects on the GPS signal. These effects are a function of frequency and the number of free electrons encountered by the signal. A definition and an expression for calculating each effect are given. The magnitude of each will be evaluated using the L₁ frequency of 1575.42 MHz and a “worst-case” TEC_{path} value of 10¹⁹ electrons/m². The higher L₁ frequency is used for this analysis because these effects impact single-frequency users at this frequency. The lower L₂ frequency is available only to dual-frequency users who correct for ionospheric effects directly. This analysis using the worst-case TEC will yield errors much larger than those typically expected, but will be useful for highlighting which effects are significant and which may be neglected. More realistic values can be obtained by using the expression given for each and substituting expected values for TEC and other parameters for a given SRP-4 launch scenario.

Refractive Index: The observable ionospheric effects on a radiowave can be traced to the effective index of refraction, n , for this region. Due to the presence of free electrons and their interaction with the electromagnetic wave, n is not equal to 1. Allnutt [9] gives the following expression for the index of refraction:

$$n = \sqrt{1 - \frac{f_p^2}{f^2}} \quad (4)$$

where: f_p = the plasma frequency in MHz,
 f = the radiowave frequency,

and $f_p = 8.9788 \times 10^{-6} \times \sqrt{N}$ Megahertz. (5)

where: N = number of electrons/m³.

Parkinson and Spilker [3] give the following first-order approximation based on the first two terms of a Taylor series expansion of (4):

$$n = 1 - \frac{f_p^2}{2f^2} \quad (6)$$

where: f_p = the plasma frequency,
 f = the radiowave frequency.

It is good to within 1% and is valid when $(f_p/f)^2 \ll 1$. For cases where f approaches f_p , equation (4) should be used. Since plasma frequency is a function of electron density, equations (4) and (6) demonstrate that refractive index is a function of signal frequency and the number of free electrons encountered.

Critical Frequency: The existence of a variable index of refraction leads to the discussion of critical frequency. Based on critical reflection theory, in cases where the signal frequency is less than the plasma frequency, equation (4) yields an imaginary refractive index, and the radiowave is completely reflected. For cases where the signal frequency is greater than the plasma frequency, a positive real refractive index results, and some part of the radiowave passes through the medium. The plasma frequency is often called the critical frequency, f_c , and it is the boundary frequency below which n becomes imaginary and no signal passes through. This critical frequency may be calculated [9] as:

$$f_c = 8.9788 * 10^{-6} \times \sqrt{N} \text{ Megahertz} \quad (7)$$

where: N = the number of electrons/m³.

Given that ionization is the mechanism producing these free electrons, equation (7) indicates that the more ionized the region is, the higher the frequency needs to be to penetrate it. Thus, the ability of a signal to transit the ionosphere is a function of both frequency and electron density. In this instance, using the lower GPS frequency of 1227.60 MHz for a worst-case analysis, equation (7) yields an electron density, N , of 1.87×10^{16} electrons/m³ that would have to be exceeded for the signal to be totally reflected. This limiting value is four orders of magnitude above typical values of electron density, which are in the 10^{12} electrons/m³ range [9]. Therefore, the frequency of the L-band GPS signal is higher than f_c for the ionosphere, and, obviously, at least part of the signal does indeed propagate through this region. An evaluation of the signal's characteristics when it emerges will now be completed in terms of specific effects as functions of frequency and TEC.

Faraday Rotation: One significant effect is a rotation of the polarization vector caused by the ionosphere. The magnitude of this rotation can be calculated using [9]:

$$\phi = \frac{2.36 * 10^4}{f^2} \times B_{av} \times TEC_{path} \quad (8)$$

where: ϕ = rotation of the polarization vector in radians,

$$\begin{aligned}
B_{av} &= \text{earth's average magnetic field, Wb/m}^2 \text{ (Teslas),} \\
f &= \text{frequency, Hz,} \\
TEC_{path} &= \text{Total Electron Content along the path in electrons/m}^2.
\end{aligned}$$

The magnetic field of the earth as a function of height and magnetic latitude may be approximated by [10]:

$$B = \frac{\sqrt{(1 + \sin^2 \lambda)} B_0}{R^3} \quad (9)$$

where:

$$\begin{aligned}
B &= \text{earth's average magnetic field, Wb/m}^2 \text{ (Teslas),} \\
\lambda &= \text{magnetic latitude in degrees,} \\
B_0 &= \text{magnetic field at the magnetic equator at the earth's surface in} \\
&\quad \text{Wb/m}^2 \text{ (Teslas),} \\
R &= \text{radial distance measured in units of earth radii.}
\end{aligned}$$

This equation may be used to determine a B_{av} encountered by a particular signal. The B_0 is typically taken to be 0.30 gauss which is 3×10^{-5} Wb/m² (Teslas). As may be seen from equation (9) the field strength decreases with height and increases with magnetic latitude. Therefore, magnetic latitude and ionospheric height for a particular signal are critical for getting an actual B_{av} and therefore an actual Faraday rotation. However, a maximum B_{av} that exceeds the actual ionospheric value may be calculated for $R=1$ and $\lambda=90$ degrees. Using a TEC_{path} of 10^{19} electrons/m², $f=1575.42$ MHz, and a worst-case $B_{av}=4.24 \times 10^{-5}$ Wb/m² (Teslas), the maximum Faraday rotation of a GPS signal is approximately 4 radians. Clearly there exist combinations that produce rotations near 90 degrees which would cause significant difficulties when using a linearly polarized signal. For the circularly polarized GPS signal, the effect is negligible.

Ionospheric Group Delay: Radiowaves transiting the ionosphere also experience a group delay. Allnutt [9] and Parkinson and Spilker [3] calculate the signal group delay to be:

$$\Delta t = \frac{40.3}{cf^2} \times TEC_{path} \quad (10)$$

where:

$$\begin{aligned}
\Delta t &= \text{signal time delay caused by ionosphere in seconds,} \\
c &= \text{speed of light in a vacuum in m/s,} \\
f &= \text{frequency of signal in Hz,} \\
TEC_{path} &= \text{Total Electron Count along path in electrons/m}^2.
\end{aligned}$$

For a TEC_{path} of 10^{19} electrons/m², $f=1575.42$ MHz, and $c=3.0 \times 10^8$ m/s, the time delay due to the ionosphere calculated using equation (10) is 5.412×10^{-7} seconds. When solving for a

distance to a satellite, this time delay turns into an absolute range error of 162.37 m using equation (1). Fortunately, this extreme value is partly compensated for by the ionospheric model and correction described later.

Ionospheric Carrier Phase Advance: As noted by Allnutt, "Delaying the arrival of a signal is similar to advancing the phase of the received signal; i.e. it appears to have traveled further than it actually has [9]." Both Allnutt [9] and Parkinson and Spilker [3] calculate the phase advance in cycles to be:

$$\Delta\phi = \frac{(1.34 \times 10^{-7})}{f} \times TEC_{path} \quad (11)$$

where: $\Delta\phi$ = signal phase advance in cycles,
 f = frequency in Hz,
 TEC_{path} = Total Electron Content along the path in electrons/m².

For a TEC_{path} of 10^{19} electrons/m² and $f = 1575.42$ MHz, the phase advance of the signal calculated using equation (11) is 850.57 cycles. When solving for the distance to a satellite, this phase advance equates to a time delay of 5.40×10^{-7} seconds at a frequency of 1575.42 MHz, which in turn becomes an absolute range error of 161.97 m using equation (1). It may be seen that this different approach yields the same range error as that found from equation (9), as would be expected. Again, this extreme value is partly compensated for by the model to be described later.

Ionospheric Doppler Shift: In addition to measuring position, GPS is used to determine a user's velocity. The geometric Doppler shift is measured on the signal from each satellite in order to calculate this velocity. As seen earlier, the ionosphere causes a phase advance and "because frequency is simply the time derivative of phase, an additional contribution to geometric Doppler shift results because of changing TEC. This additional frequency shift is generally small compared with the normal geometric Doppler shift [3]." It can be computed by:

$$f_D = \frac{1.34 * 10^{-7}}{f} \times \frac{d}{dt} (TEC_{path}) \quad (12)$$

where: f_d = ionospheric Doppler frequency shift in Hz,
 f = frequency in Hz,
 TEC_{path} = Total Electron Content in electrons/m².

Parkinson and Spilker report that the "upper limit to the rate of change of TEC for a stationary user is approximately 0.1×10^{16} (electrons/m²) per second. This value yields an additional

frequency shift of 0.085 Hz at L_1 , which would not be significant, compared with a typical required receiver carrier tracking loop bandwidth of at least a few Hz. The value of 0.085 Hz at L_1 corresponds to 1.6 cm/s of range-rate error [3]." This small velocity error is insignificant for most GPS applications, but could possibly be a factor for some high-accuracy scientific uses.

Angular Refraction: The effective refractive index of the ionosphere causes a bending of the radiowave path. This means the signal has traveled a longer distance than the straight-line "ray" path and this violates one of the assumptions discussed earlier and results in a position measurement error. The observed effect of this bending is a change in angle-of-arrival of the signal at the user's receiver antenna. This "produces an apparent higher elevation angle than the geometric elevation [3]," but most GPS receivers use omni-directional antennas, so angle of arrival is not a concern for antenna pointing and tracking. Additionally, as seen in Figure 10, these angle-of-arrival errors are small at all elevations, and very small at higher elevations. Most receivers "mask" out satellites below 5 or 10 degrees elevation in order to minimize other effects due to the troposphere, multipath, etc. that can cause large errors at low elevations. This also helps to minimize the impact of this effect, and angular refraction results in small distance

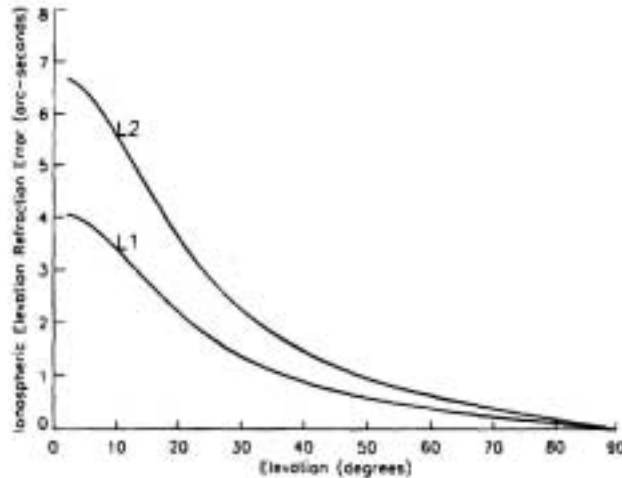


Fig. 10. Ionospheric refraction vs. elevation for both L_1 and L_2 . From Fig. 3. of reference [3].

measurement errors. Allnutt [9] confirms this conclusion by stating that for frequencies above 1 GHz, like GPS, angle-of-arrival problems induced by the ionosphere are insignificant. Parkinson and Spilker [3] also note that azimuthal gradients in TEC exist, but that they are generally smaller than vertical gradients and therefore angular refraction is a weaker function of azimuth angle than elevation angle and should be negligible.

Distortion of Pulse Waveforms: "The GPS signals consist of spread spectrum pseudorandom noise, having bandwidths of approximately 2 MHz and 20 MHz [3]" for the codes on L_1 and L_2 respectively. The ionosphere causes a dispersion of spread spectrum signals as they pass through

which manifests itself as a difference in pulse arrival time across a bandwidth, Δf . This dispersion is calculated [9], [3] using:

$$\Delta t = \left[\frac{(80.6 \times \Delta f)}{cf^3} \right] \times TEC_{path} \quad (13)$$

where:

- Δt = difference in pulse arrival time across a bandwidth in seconds,
- Δf = bandwidth in Hz,
- c = speed of light in a vacuum in m/s,
- f = frequency of signal in Hz,
- TEC_{path} = Total Electron Content along the path in electrons/m².

Using the L₁ bandwidth of 2 MHz, $c=3.0 \times 10^8$ m/s, the L₁ frequency of $f=1575.42$ MHz, and a TEC_{path} of 10^{19} electrons/m², the dispersion is calculated using equation (13) to be 2.12×10^{-13} seconds. This is a small number and does not represent a significant impact on the signal. Additionally, for L₂ "The dispersion across the 20-MHz GPS bandwidth is normally small and can be ignored [3]."

Ionospheric Scintillation: Allnutt describes scintillation of a radio signal as "a relatively rapid fluctuation of the signal about a mean level that is either constant or changing much more slowly than the scintillations themselves [9]." Parkinson and Spilker note that "irregularities in the Earth's ionosphere produce both diffraction and refraction effects, causing short-term signal fading [3]." As with all scintillation, in addition to fades, signal enhancements also occur. Since GPS position determination relies on the downlink, and there is no transponder involved, there is no concern about degradation of signals on other channels due to increases in power. Generally, GPS receivers are designed to operate and maintain lock above some minimum threshold signal level so the enhancements neither help nor hurt. The real concern is with fades deep enough to cause a loss of signal lock at the receiver. Figure 11 shows the regions of the earth over which scintillations are most likely and most severe. Maximum scintillations occur near and within the short-dashed lines, and the long-dashed line indicates the Earth's magnetic equator. In this region, amplitude fades >25dB have been recorded and may last minutes or remain for periods up to several hours. Frequency and severity of scintillations are a function of both time of day and solar cycle. Worst-case values are generally encountered between one hour after sunset and

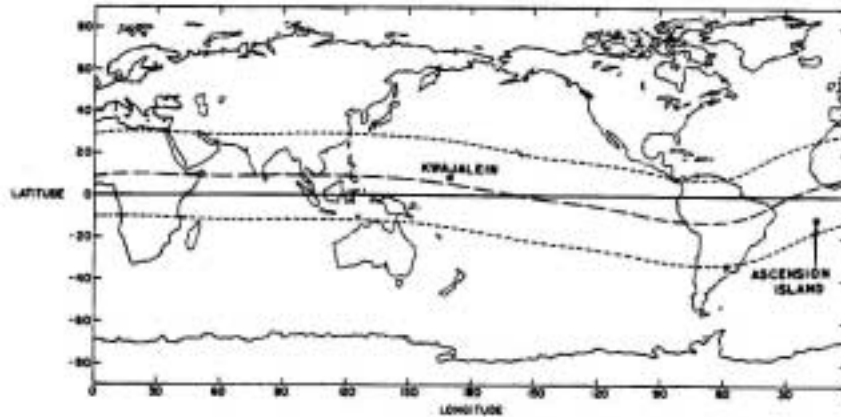


Fig. 11. Geographic regions of maximum L-band ionospheric scintillation. From Fig. 6 of reference [3].

midnight and during periods of solar maximum [3]. Figure 12 is one representation of these dependencies. There are also scintillations in the auroral and polar cap regions that are associated with magnetic storm activity, but these are typically not as severe as those measured in the near-equatorial regions indicated in figure 11. They can, however, last "for many hours, even days, and are not limited to the local late evening hours, as the near equatorial scintillation effects are [3]." This is one area that warrants further research for the SRP-4 case. The occurrence of both amplitude and phase scintillations can be represented in a statistical sense, but long-term data on GPS scintillations appears to be scant. In addition, Allnutt states that "the concept of annual statistics is of dubious merit for ionospheric phenomena...due to the cyclic dependence of ionospheric scintillation intensity on the equinoxes and the 11-year sun-spot cycle [9]." While both amplitude and phase scintillations can be severe enough to cause loss of lock in the receiver, these occurrences are infrequent. Of course, if they happen during a launch, they become problematic.

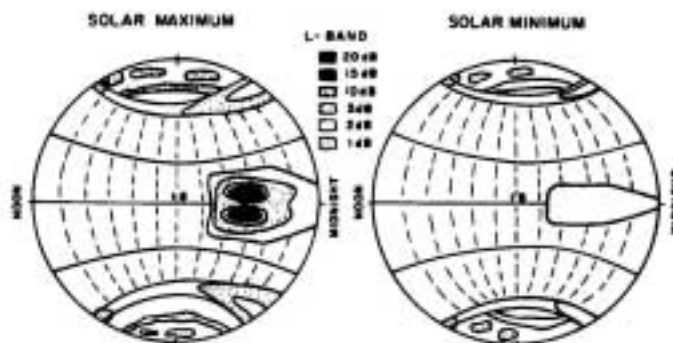


Fig. 12. Worst case L-band ionospheric scintillation fading depths. From Fig. 5 of reference [3].

APPROACHES TO MINIMIZE IONOSPHERIC EFFECTS

A number of design and system operation choices have been made in order to minimize the detrimental effects of the ionosphere on GPS signals. The design choice to transmit a right-hand circularly polarized signal eliminates the need to constantly align an antenna to couple efficiently with a linearly polarized signal. In practice, most user receiver antennas are linearly polarized to eliminate the bore sighting required to align an optimum circularly polarized receive antenna to couple with the incoming signal. This choice of antenna results in a 3-dB loss because only one-half of the potential signal energy is being received. "The up to 3-dB loss between transmitted circular polarization and receiver nearly-linear polarization is a necessary price GPS users pay for antenna maneuverability and simplicity. The transmitted signal levels from the GPS satellites were designed to provide adequate signal strength for users with linearly polarized antennas [3]."

To minimize the various error contributions to GPS position and velocity solutions, including ionospheric effects, the path length through the atmosphere is shortened by using only satellites above an elevation of 5-10 degrees. The orbital constellation was designed to account for this limitation and, barring excessive satellite outages, should still allow for the minimum of four satellites with acceptable solution geometry for a user anywhere on the globe.

Since these ionospheric effects can be predicted using the expressions in the previous section, given an accurate worldwide model of ionospheric TEC values, even a single-frequency user should be able to correct his solution for these effects. Unfortunately, the 1σ day-to-day variability of the TEC from the monthly mean value is approximately 20-25% [3], making it very difficult to model effectively. The more accurate ionospheric models are too cumbersome to use directly. In practice, GPS uses a fairly simple model that will "correct for approximately 50% of the ionospheric range error [3]." The simplicity of this single-frequency receiver model was driven by the fact that the broadcast message only had room for eight parameters and the parameters could only be updated once a day, if that, when the control station uploaded the navigation message to the satellite [3]. This is a significant limitation due to the variability of the ionosphere on a time scale less than one day, and the "compensation" degrades as the coefficients get older. It models the monthly median behavior and uses a "half-cosine" form to model the diurnal variation about that monthly median behavior. This model takes the form of [3]:

$$T_{iono} = F \times \left\{ DC + A \cos \left[\frac{2\pi(t - \phi)}{P} \right] \right\} \quad (14)$$

where:

- F = slant factor to account for slant path length,
- DC = the DC offset term,
- A = amplitude,
- ϕ = phase of the maximum with respect to local noon,
- P = period.

Since F is calculated by the receiver based on location, there are four other parameters in the model that vary daily and only eight coefficients to represent them on a worldwide basis. Therefore, the decision was made to set the two that vary the least, DC and ϕ , to a constant value, and use the broadcast coefficients to cover the two that vary the most, A and P. The receiver then uses these coefficients in an algorithm to mathematically model the current behavior of the earth's ionosphere [11].

The Ionospheric Correction Algorithm used by the receiver, as reported by Parkinson and Spilker [3], is outlined below. User inputs to the model are the user's approximate geodetic latitude Φ_u , longitude λ_u , elevation angle E , and azimuth A to each satellite. The coefficients α_n and β_n are the polynomial coefficients that represent amplitude and period as described above, and are uploaded to the satellite by the ground segment. They are transmitted to the user as part of the navigation message broadcast by each satellite, and the ionospheric time delay is estimated by a single-frequency receiver as follows [3]:

- 1) Calculate the Earth-centered angle, ψ

$$\psi = \frac{0.0137}{(E + 0.11)} - 0.022 \quad (\text{semicircles}) \quad (15)$$

- 2) Compute the subionospheric latitude, Φ_I

$$\Phi_I = \Phi_u + \psi \cos A \quad (\text{semicircles}) \quad (16)$$

If $\Phi_I > 0.416$, then $\Phi_I = 0.416$. If $\Phi_I < -0.416$, then $\Phi_I = -0.416$.

- 3) Compute the subionospheric longitude, λ_I

$$\lambda_I = \lambda_u + \left(\frac{\psi \sin A}{\cos \Phi_I} \right) \quad (\text{semicircles}) \quad (17)$$

- 4) Find the geomagnetic latitude, Φ_m , of the subionospheric location looking toward each GPS satellite.

$$\Phi_m = \Phi_I + 0.064 \cos(\lambda_I - 1.617) \quad (\text{semicircles}) \quad (18)$$

- 5) Find the local time, t , at the subionospheric point

$$t = 4.32 \times 10^4 \lambda_I + \text{GPS time} \quad (\text{seconds}) \quad (19)$$

If $t > 86,400$, use $t = t - 86400$. If $t < 0$, add 86400.

6) To convert to slant time delay, compute the slant factor, F

$$F = 1.0 + 16.0 \times (0.53 - E)^3 \quad (20)$$

7) Compute the ionospheric time delay T_{iono} by first computing x

$$x = \frac{2\pi(t - 50400)}{\sum_{n=0}^3 \beta_n \Phi_m^n} \quad (21)$$

If $|x| > 1.57$, then

$$T_{iono} = F \times (5 \times 10^{-9}) \quad (22)$$

Otherwise,

$$T_{iono} = F \times \left[5 \times 10^{-9} + \sum_{n=0}^3 \alpha_n \Phi_m^n \times \left(1 - \frac{x^2}{2} + \frac{x^4}{24} \right) \right] \quad (23)$$

Again, this simple model typically corrects between 50 and 70 percent of the error due to ionospheric effects.

The two-frequency user, as alluded to earlier, can derive an actual TEC value and TEC rate of change along the signal path. This is made possible by the very fact that ionospheric effects are a function of frequency as illustrated earlier. Once an accurate TEC value is derived, all the associated effects can be calculated and corrected for. One approach to accomplishing this is to consider an error in range as an error in time as described by Allnutt [9]. Since the L_1 and L_2 signals have a common timing reference in the modulation applied to them, the relative error in time can be measured and equation (10) or equation (11) can be rearranged to yield the difference in time error as:

$$\delta t = \frac{40.3}{c} \times \left(\frac{1}{f_2^2} - \frac{1}{f_1^2} \right) \times TEC_{path} \quad (24)$$

where: δt = difference in time error in seconds,
 c = speed of light in a vacuum in m/s,

$$\begin{aligned}
f_2 &= L_2 \text{ frequency in Hz,} \\
f_1 &= L_1 \text{ frequency in Hz,} \\
TEC_{path} &= \text{Total Electron Content along the path in electrons/m}^2.
\end{aligned}$$

Once this difference in time error is determined, it may be used in the following to calculate TEC_{path} [9]:

$$TEC_{path} = \frac{\delta t \times c}{40.3} \times \frac{f_1^2 f_2^2}{f_1^2 - f_2^2} \quad (25)$$

where: TEC_{path} = Total Electron Content along the path in electrons/m²,
 δt = difference in time error in seconds,
 c = speed of light in a vacuum in m/s,
 f_2 = L₂ frequency in Hz,
 f_1 = L₁ frequency in Hz.

A final approach for ameliorating the effects of the ionosphere is a technique used to compensate for many other error sources as well. The discussion thus far has centered on the process a stand-alone receiver uses to calculate a position or velocity solution. As discussed earlier, differential GPS can be used to eliminate many types of error and ionospheric errors are one. The details of what accuracy can be expected from differential GPS will be discussed in a later section.

OVERALL IMPACT ON PSEUDORANGE MEASUREMENT

Clearly there are a number of ionospheric effects that impact the performance of GPS. Thanks to the design choice to use circular polarization, Faraday rotation is not a concern, but the omni-directional antennas used on most receivers have to deal with a 3 dB loss as the linear antenna couples with only half of the circularly polarized signal. Fortunately the transmitted signal strength was designed with this in mind. The analysis of ionospheric group delay and phase advance looks at the same error from two different perspectives. This is the most significant effect in terms of pseudorange measurement accuracy and for a "worst-case" ionosphere without the benefit of the correction algorithm, this error is on the order of 162 m for the distance measurement to each of the reference satellites. With the correction applied, it can range from .05 m to 81 m depending on how active the ionosphere is and the elevation angles to the satellites used in the solution. The additional Doppler shift, the angular refraction effects, and the signal distortion effects all prove to be negligible for even a maximum slant path TEC. Parkinson and Spilker [3] report that a typical value for temperate zones, after correction by the algorithm, is 4.0 m. As will be discussed later, there are likely ionospheric effects in the auroral region that warrant further investigation. Indeed, while "the ionosphere is reasonably well-behaved and stable in the temperate zones; near the equator or magnetic poles it can fluctuate

considerably [3]."

A.3.1.4 TROPOSPHERIC ERROR

Tropospheric errors are functions of temperature, humidity and pressure and range from .3 m to 3.0 m. For frequencies below 15 GHz, this part of the atmosphere is non-dispersive, or not a function of frequency [4]. Therefore these errors cannot be corrected, as can ionospheric errors, by observing the different effects at the L_1 and L_2 frequencies. For GPS, that leaves models as the approach for predicting effects. Unfortunately, the troposphere is extremely variable, especially in what is known as the "wet component" and is very difficult to model. Fortunately, the tropospheric errors are smaller in magnitude than the ionospheric errors, and for most users, "a simple model should be effectively accurate to about 1 m or better [3]." Again, the table value of 0.7 m is based on time-averaged effects [8]. As will be discussed in more depth later in this paper, one area of concern for SRP-4 is that the high-latitude troposphere causes significantly different effects than the temperate troposphere on which these time-averaged effects are likely based.

A.3.1.5 RECEIVER ERROR

The receiver tracking loops also induce pseudorange measurement errors. Kaplan [4] reports that the dominant sources of pseudorange error are thermal noise jitter and dynamic stress error. Smaller errors are attributed to code hardware and software resolution and oscillator stability. Typical 1σ errors for a modern receiver are reported by Kaplan [4] to be 1.5 m for C/A-code. Parkinson and Spilker [3] report a smaller 1σ error of 0.5 m as typical. The second source appears to be based on more recent information, so that value is used in the table of predicted PR errors.

A.3.1.6 MULTIPATH ERROR

Multipath errors are due to signals arriving at the receiver via multiple paths. Figure 13 demonstrates how this might occur. These multiple paths are generated by reflections off of the Earth, buildings, fins of the rocket, etc. Under worst-case conditions, multipath may be bad enough to cause a receiver to lose lock on the signal [4], although more often it causes a PR measurement error. "The degradation of the pseudoranges is caused by the distortion of the correlation peak by the presence of the indirect signal...Distortion of the correlation peak causes the zero crossing of the early-late curve to be shifted and thus the receiver will determine an erroneous pseudorange [4]." The severity of the multipath error is a function of amplitude, phase, phase rate, and delay of the indirect signal with respect to the direct signal. In fact, as the path delay of a reflected signal increases with respect to the path delay of the direct signal, the error will grow until a maximum is attained and then will decrease again to zero, and then grow again as the distance increases further. A more detailed discussion of this is found in [4].

Finally, the magnitude of the multipath error will depend greatly upon the antenna installation on

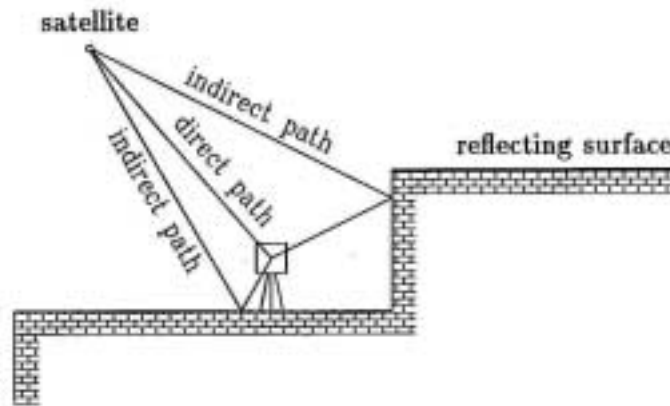


Fig. 13. Multipath effect. From Fig. 6.5. of reference [5].

the rocket and attitude of the rocket with respect to the incoming signals. Once the rocket is away from the launch facility, it is difficult to imagine scenarios where a reflected signal will reach the rocket antenna with much signal strength so this error may be small. Without details of the actual installation, a typical value will be used. Kaplan [4] reports a typical value to be 2.5 m and Parkinson and Spilker [3] use 1.4 m. Based on the assumption that the only likely reflector will be the ground once the rocket is aloft, the smaller value of 1.4 m will be used in the table.

A.3.2 BOTTOM LINE: WHAT ACCURACY CAN BE EXPECTED?

The following table gives the predicted PR errors for the flight GPS system when operated as a stand-alone unit, without differential corrections. The table entries are "typical" 1σ values derived from the analysis in the preceding section.

Table 1: Predicted PR errors for stand-alone C/A-code GPS

Error Source	Typical Value (m)
Satellite clock stability	2.1
Ephemeris prediction error	4.2
Ionospheric delay	4.0
Tropospheric delay	0.7
Receiver noise and resolution	0.5
Multipath	1.4
System UERE (rms)	6.38

The errors from the various sources are assumed to be uncorrelated and of zero mean [3] and are combined in a root-sum-square (rss) fashion to get an overall PR error. As stated in section 1.4,

this PR error is commonly referred to as the user-equivalent-range-error, or UERE, and may be treated as a root-mean-square (rms) value due to the manner in which it is derived [3], [4]. Since these are statistical measurements, the actual error contributions for a given launch could be better or worse than those represented in the table.

As described earlier, the position error is not only a function of the UERE, but also a function of the geometry, or PDOP. To estimate the overall error in the position solution for a given choice of satellites, the UERE is multiplied by the PDOP for that set of satellites:

$$Error_{1\sigma} = UERE_{rms} \times PDOP \quad (26)$$

where: $Error_{1\sigma}$ = the one-sigma GPS position error in meters,
 $UERE_{rms}$ = the user-equivalent-range-error in meters,
 PDOP = the position dilution of precision.

This error is based on the "rms" UERE and is considered to be a 1σ error corresponding to a probability of 61-68% [7]. A 2σ error corresponding to 95% probability [8] is calculated as:

$$Error_{2\sigma} = 2 \times UERE_{rms} \times PDOP \quad (27)$$

where: $Error_{2\sigma}$ = the two-sigma GPS position error in meters,
 $UERE_{rms}$ = the user-equivalent-range-error in meters,
 PDOP = the position dilution of precision.

Substituting the table UERE value and the worldwide median PDOP of 2.5 [3] into equation (26) yields a predicted 1σ error of 16.0 meters. The 2σ error from equation (27) with the same inputs is 31.9 meters. These are "typical" accuracies with the associated probabilities described above.

To obtain a more accurate prediction, given a specific launch date and time, the predicted TEC_{path} can be substituted into equation (10) or equation (11) to model the ionospheric delay. The error calculated using this equation is typically reduced 50 to 70 percent by the receiver's internal ionospheric model. After making this adjustment, this new value can be used to generate a revised UERE. In addition to using an updated UERE, a PDOP for the specific launch time can be calculated using a software package as discussed earlier and used in place of the worldwide median value. This and the new UERE can then be substituted into equation (26) or equation (27) to find a more accurate expected position error.

A.4 PREDICTED PERFORMANCE OF DIFFERENTIAL C/A-CODE GPS

Differential GPS systems take advantage of the fact that many of the errors are slowly changing with time and are strongly correlated over distance. This allows the reference receiver to generate corrections and apply them before the corrections are "out of date". It also allows the

corrections to be applied to the solution for a mobile receiver many kilometers away. Uncorrelated errors change with both time and distance and cannot be corrected using differential GPS. The following analysis evaluates each of the significant error sources with respect to the cause of the effect, how correlated it is with time and distance, and what the residual error is expected to be following differential correction. A summary of the individual expected error contributions is included in Table 2 at the end of this section and is followed by an assessment of the predicted position accuracy for the differential system.

A.4.1 PREDICTED PSEUDO-RANGE ERRORS

The following evaluates the “typical” error components that remain after differential corrections are applied. It takes into account the reported magnitude as well as the degree to which each error is correlated with time and distance.

A.4.1.1 SATELLITE CLOCK ERROR

Clock errors decorrelate with time, but not with distance, and are typically correlated for up to 5 minutes [8]. Therefore, a significant part of this error should be correctable using DGPS techniques as long as the correction is applied promptly. The residual 1σ error after correction is approximately 0.7 meters [8].

A.4.1.2 EPHEMERIS ERROR

Ephemeris errors decorrelate with distance, not with time, but are typically only 0.05 m different over 100km [8]. Therefore, as long as the mobile receiver is within 100 km horizontally from the reference receiver, we can expect small residual errors. Figure 14 illustrates the expected residual position error given a separation distance between the reference

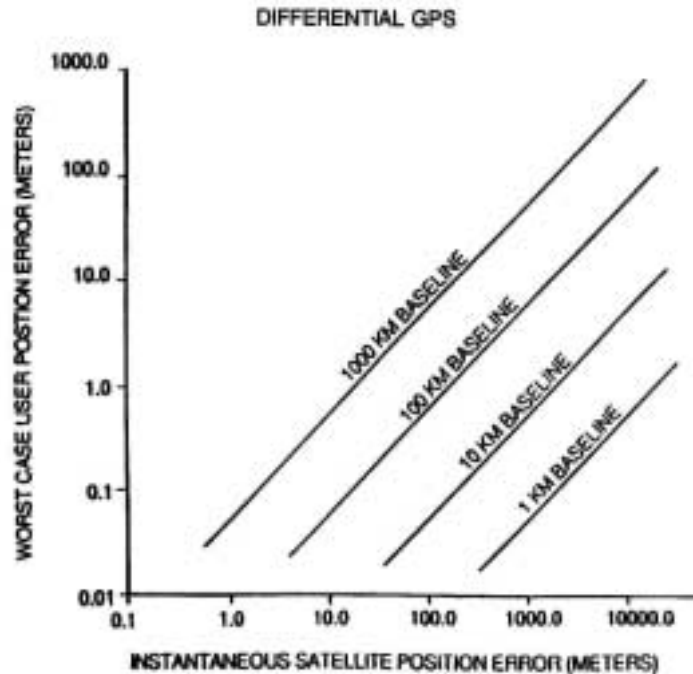


Fig. 14. Worst-case differential GPS errors due to satellite ephemeris error for various distances from reference to user. from Fig. 1.6 of reference [8].

station and the mobile receiver. Parkinson and Spilker conservatively estimate a typical expected value to be on the order of 0.5 meters [8].

A.4.1.3 IONOSPHERIC ERROR

Ionospheric errors decorrelate with time and distance, but "As long as *both or neither* the user and reference station make a dual-frequency correction, the impact on DGPS should be errors of less than 1 m for separations of less than 100 km [8]." Differential corrections for ionospheric effects are sometimes inaccurate when the signal path to the reference receiver cuts through part of the ionosphere with a substantially different TEC than the TEC found along the path to the mobile receiver. They also grow "old" and lose accuracy as the time period between calculation and application increases, especially when the ionosphere is rapidly changing. Spilker and Parkinson [8] report that the geographical decorrelation is less than 0.2 m per 100 km horizontal separation. Combining this with the time decorrelation effects, they give a conservative residual error of 0.5 m.

A.4.1.4 TROPOSPHERIC ERROR

Tropospheric errors are also correlated with time and distance, but are much smaller than ionospheric errors. Up to 90% of the tropospheric error is removed by a simple model dependent

only on elevation, and the differential system is only useful for removing the remaining small correlated error [8]. One area of concern for SRP-4 is due to the fact that the differential station is located at ground level and is below the troposphere, while the rocket is above much of the troposphere for a critical portion of its flight. This may result in inaccurate corrections being applied for tropospheric error. Since the tropospheric error is small, the resulting inaccuracy should be small. The typical 1σ error not correctable by DGPS is reported to be 0.5 meters [8].

A.4.1.5 RECEIVER AND MULTIPATH ERRORS

These errors are grouped together because each decorrelates over very short distances and very short time periods [8]. Therefore these errors cannot be effectively corrected using differential methods. Instead, it is very important that they be minimized at the reference station, because any correction generated to compensate for these effects, that is then applied to the pseudoranges from the mobile receiver, introduces error instead of removing it. Assuming that no artificial errors are included in the differential correction, the typical values for these errors are the same as those found in the stand-alone system.

A.4.2 BOTTOM LINE: WHAT ACCURACY CAN BE EXPECTED?

Substituting this UERE value and the worldwide median PDOP of 2.5 into equation (26) yields a predicted 1σ error of 4.7 meters. The 2σ error from equation (27) using the same inputs is 9.3 meters. These are “typical” accuracies with the associated probabilities discussed earlier for 1σ and 2σ values. It should be noted that any "artificial errors" introduced by the reference station

Table 2: Predicted PR errors for differential C/A-code GPS [8]

Error Source	Typical Value (m)
Satellite clock stability	0.7
Ephemeris prediction error	0.5
Ionospheric delay	0.5
Tropospheric delay	0.5
Receiver noise and resolution	0.5
Multipath	1.4
System UERE (rms)	1.86

forming a correction for decorrelated errors will result in a larger UERE. As in the stand-alone case, the PDOP for the actual flight time should be calculated using a software package and this PDOP should be substituted into equations (26) and (27) to generate more accurate predictions.

A.5 AREAS IN NEED OF FURTHER INVESTIGATION

The error assessment conducted in this paper quantifies the “typical” errors in positional accuracy expected from a single-frequency C/A-code GPS receiver. A number of areas remain to be evaluated in order to improve this evaluation for SRP-4. These areas should focus on those aspects of the SRP-4 environment that are not typical and pose the greatest likelihood of introducing additional pseudorange errors.

A.5.1 HIGH-LATITUDE IONOSPHERIC PHENOMENA

Since SRP-4 will be launched from Poker Flat Research Range at 65 degrees north latitude, high-latitude ionospheric effects should be more thoroughly investigated. The presence of aurora alone is evidence that the ionosphere at high-latitudes is indeed different than that found in more temperate regions. Since ionospheric effects are often the largest contributor to pseudorange error, those mechanisms that cause regional and localized changes in ionization should be evaluated for their practical impact on the expected positional accuracy.

A.5.2 HIGH-LATITUDE TROPOSPHERIC PHENOMENA

A second area in need of further investigation is the existence, or lack thereof, of high-latitude tropospheric effects that are different than those encountered in the mid to lower latitude regions. Most tropospheric analysis to date has centered on the temperate regions, and not on high-latitudes. Indeed, the analysis in this paper has depended on “typical” values that most likely are derived from temperate region data. While impact on GPS by the troposphere has been shown to be much less than that due to the ionosphere, tropospheric effects seem to be the most likely error to depend on latitude, after ionospheric effects, and should be further investigated to evaluate their practical effect on expected positional accuracy

A.5.3 HIGH-LATITUDE PDOP CONSIDERATIONS

A third area that warrants further investigation is the effect of high-latitude operation on PDOP. Since positional accuracy depends on both pseudorange errors and the geometry indicated by dilution of precision, poor PDOP is a concern. Specifically, the relationship between the 55 degree inclined GPS satellite orbits and the view of these satellites from Poker Flat should be evaluated. This is not likely to be a problem since GPS satellites at an altitude of approximately 26,560 km can probably be seen over the north geographic pole despite their lower inclination. If this is the case, PDOP values should be good, and error correspondingly small. Before a final determination of expected accuracy is completed, this evaluation should be made.

A.5.4 RELATIVISTIC EFFECTS

Finally, the fact that a sounding rocket travels at high velocity and to high altitude raises the possibility of position errors due to relativistic effects. Since GPS relies on the accurate time-tagging of signals and the precise measurement of transit time, the effects of general and special relativity on both the satellite and receiver clocks are important. These effects from the perspective of a ground-based user are treated in depth in [3] and [4]. To summarize, the satellite atomic clocks are offset before launch so that a ground-based receiver perceives the frequency to be correct, and most modern receivers correct for the other effects. In the case of SRP-4, however, the velocity of the receiver with respect to the satellites can be much higher than that of a ground-based user and the gravitational field will change as well. This operating envelope results in additional special and general relativity effects, not accounted for by the corrections mentioned above. Parkinson and Spilker note that some of these effects can cancel out to varying degrees, but that "These effects can be significant if the user is another satellite in orbit [3]." While sounding rockets do not achieve orbit, their altitudes and velocities are very different from ground-based users. The magnitude of these effects should be further evaluated to assess their impact on positioning accuracy.

A.6.0 POSSIBLE ENHANCEMENTS TO IMPROVE FUTURE POSITIONING ACCURACY

While the current design should easily meet any foreseeable accuracy requirements for range safety, trajectory analysis, or payload recovery, there may be needs for higher accuracy in the future driven by new payload requirements. The following are a few areas that might be evaluated for future designs if there is a validated requirement.

A.6.1 DUAL-FREQUENCY RECEIVER

One large improvement may be achieved by using a dual-frequency receiver. Having two separated frequencies allows the receiver to correct for the ionospheric effects directly, thereby eliminating the largest single contribution to the UERE. One method by which this might be accomplished was discussed earlier in this paper. Unfortunately, the U.S. government currently encrypts the second GPS frequency and it is available only to authorized users with a verifiable need. There is some discussion of a second civilian frequency to be implemented on future generations of GPS satellites. If this comes about, it would significantly improve positioning accuracy in applications such as this.

A.6.2 "HIGH-DYNAMIC" RECEIVER

The foregoing analysis assumes that the receiver and associated hardware function as

designed. One area of concern is the ability of the GPS receiver to lock-up and track during periods of high acceleration and high “jerk”. If it is determined after the SRP-4 launch that data was lost during critical periods due to an inability to lock up and track under these conditions, it may be fruitful to evaluate a “high-dynamics” receiver. The NASA-Wallops system uses a receiver recently developed for missile applications that is advertised to be less susceptible to these events. This type of receiver is currently cost-prohibitive for the SRP-4 payload, but the price may come down in time for future missions.

A.6.3 CARRIER-PHASE METHODS

A more extreme redesign of this system could move from code-based techniques to carrier-phase methods. Instead of evaluating the pseudorandom noise code to develop a time since transmission, these systems work with the phase of the carrier and evaluate phase shift to calculate transit time. This type of system has been proven capable of providing sub-meter accuracy. Difficulties in implementing this type of system include the more complex software programming required and the susceptibility of the hardware to the harsh environment of a sounding rocket flight. While these are not insurmountable, the added complexity would only be justified if there was a verifiable requirement for extreme accuracy. One possible driver for this change would be the desire to determine attitude, which relies almost entirely on this method.

A.7.0 CONCLUSIONS

The SRP-4 flight GPS can be expected to produce a 1σ position error of 16.0 meters when operated in a stand-alone fashion or of 4.7 meters when operated as a differentially corrected system. These error predictions are based on a number of critical assumptions. First, the ionospheric error is based on an average value and a more accurate UERE can be calculated using a Total Electron Content for the time of launch. Second, an average PDOP of 2.5 is assumed. While this is the world-wide average for the current GPS constellation, a more accurate error estimate can be made by using a software program to calculate the actual PDOP at time of launch. This PDOP can then be used with the revised UERE and a new error estimate can be calculated using equation (26). The UERE presented here necessarily used “typical” values for the other error contributions as well. It would be prudent to investigate whether additional high-latitude effects significantly change these “typical” values. Additionally, to minimize the error for SRP-4, it will be critical to minimize the multipath errors by carefully siting the differential receiver antenna. While the statistical nature of the error contributions and the other assumptions used here do not allow for a precise position error calculation, the prediction given above should be an accurate order of magnitude for the error to be expected.

A.8 REFERENCES

- [1] B. Bull, "A Real Time Differential GPS Tracking System for NASA Sounding Rockets," published by NASA Goddard Space Flight Center, Wallops Flight Facility.
- [2] B. Bull, "This is Rocket Science: Multiple Payload Tracking in Space," GPS World, pp. 22-32, Oct. 2000.
- [3] B. W. Parkinson and J. J. Spilker, Eds., *Global Positioning System: Theory and Applications I*. Washington D.C.: American Institute of Aeronautics and Astronautics, Inc., 1996.
- [4] E. D. Kaplan, *Understanding GPS: Principles and Applications*. Norwood, MA: Artech House, 1996.
- [5] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, *Global Positioning System: Theory and Practice*. New York, NY: Springer-Verlag/Wien, 1997.
- [6] B. Dahl, *The Users Guide to GPS: The Global Positioning System*. Evanston, IL: Richardsons' Marine Publishing, 1993.
- [7] F. van Diggelen, "GPS Accuracy: Lies, Damn Lies, and Statistics," GPS World, pp. 41-44, Jan. 1998.
- [8] B. W. Parkinson and J. J. Spilker, Eds., *Global Positioning System: Theory and Applications II*. Washington D.C.: American Institute of Aeronautics and Astronautics, Inc., 1996.
- [9] J. E. Allnutt, *Satellite-to-Ground Radiowave Propagation: Theory, Practice and System Impact at Frequencies Above 1 GHz*. Peter Peregrinus Ltd, on behalf of the Institution of Electrical Engineers.
- [10] W. J. Larson and J. R. Wertz, Eds., *Space Mission Analysis and Design – Third Edition*. Torrance, CA: Microcosm, 1999.
- [11] T. Logsdon, *Understanding the Navstar: GPS, GIS, and IVHS*. New York, NY: Van Nostrand Reinhold, 1995.
- [12] R. B. Langley, "GPS, the Ionosphere, and the Solar Maximum," GPS World, pp. 44-49, Jul. 2000.