2006

Precise pseudoranges obtained from combining code and dual carrier measurements in Global Positioning System receivers

Steven D. Deines
Iowa State University

Precise pseudoranges obtained from combining code and dual carrier measurements in Global Positioning System receivers

by

Steven D. Deines

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Civil Engineering (Geometronics)

Program of Study Committee:
Kandiah Jeyapalan, Major Professor
Lee Anne Willson
Say Kee Ong
John Pitt
David Carter-Lewis

Iowa State University

Ames, Iowa

2006

Table of Contents

List of Tables

List of Figures

Abstract

The Global Positioning System (GPS), which was declared operational in December 1993, has provided continuous worldwide navigational capabilities in all types of weather. GPS has a minimum of 24 satellites in its constellation with additional fully functional spare satellites, which vary about 30 in total number. This navigational service has provided three-dimensional position within 30 meters and time within 100 ns to the civil community, which typically uses the coarse acquisition (C/A) code. The military GPS user has improved position accuracy within 6 meters using the dual GPS frequencies that carry the precise (P) code. The coarse resolution of a code measurement is about 1/10 of the smallest bit length of either code, which is about 29 m on C/A or 2.9 m with P code. Also, the GPS phase measurements can be determined well within 1/100 of a cycle by both civil and military receivers, so that the equivalent wavelength portion will be less than 0.19 cm and 0.24 cm. Recent GPS technology has improved civil navigation close to the military precision limits.

To improve the precision by at least a hundred-fold, this dissertation considers a new exact linear navigation algorithm compared to the standard iterative GPS solution and also considers a new method of measurement by combining both carrier and phase measurements to improve pseudoranges within a centimeter tolerance. The GPS signal design is described in detail, and the novel techniques are derived explicitly. A simulation illustrating the novel exact solution demonstrates the greater versatility over the standard GPS iterative method, which, in some special cases, converges prematurely. Also, the GPS navigation solution is computed using both methods with actual GPS data against surveyed benchmarks. In conclusion, this dissertation: (1) derives a new exact linear GPS navigation algorithm as an alternative to the standard iterative GPS method, (2) demonstrates the standard iterative GPS navigation solution may stall prematurely in many small regions, which are dependent on the satellite configuration, before getting to the receiver's actual location, and (3) illustrates a new method that combines carrier phases with pseudorange data to obtain subcentimeter precision in the GPS pseudoranges plus improved navigation with the exact GPS algorithm.

Chapter 1 Introduction

The Global Positioning System (GPS) represents a nearly ideal dual-use technology between the civil and military communities. The original design began in 1973 when the Department of Defense decreed that GPS would be a Joint Program with a GPS Joint Program Office (GPS JPO) located at the Air Force's Space and Missile Organization in El Segundo, CA. The current GPS system was declared with Initial Operational Capability (IOC) on December 8, 1993, and it is virtually identical today as the system was proposed (Parkinson, 1996) in 1973. The satellites have expanded their functionality to support added military capabilities, and the orbits have changed from 63 degree inclinations to 55 degree inclinations for Shuttle deployment, but the GPS receivers of today would still work with the original GPS satellites. It is a major achievement that this system has kept the same design and general accuracy over three decades. All GPS signals are systematically affected by the physics of the environment due to the ionosphere, troposphere or relativity, which can be removed either by modeling or direct measurement. Still, the navigation solution is limited by inaccurate range determinations between each GPS satellite and the GPS receiver due to the noise in the code. The GPS carrier signals are 1575.42 MHz and 1227.6 MHz with a respective wavelength of 19.0 cm and 24.4 cm, respectively.

It is the purpose of this thesis to propose and demonstrate a new improvement in GPS navigation precision by combining dual phase measurements with pseudoranges as measured by a single GPS receiver in real time operation. This thesis contains a new derivation of an exact GPS Navigation solution as an alternative of the standard GPS iterative method, which can converge prematurely in some special cases. The new methods may well offer other enhancements in GPS navigation.

1.1 Principles of Global Positioning System Operation

The principal GPS navigation technique uses satellite to ground ranging information where the GPS satellites broadcast their estimated positions. The positions are obtained mathematically by trilateration from a minimum of four GPS satellites. Ranges are measured simultaneously from the satellites by each GPS receiver through

correlations of the incoming signal with a replicated signal, which obtains the time delay between transmission and reception using the receiver's internal clock. When multiplied by the speed of light, this time delay obtains the individual distances, which are called pseudoranges. These ranges have the systematic large offset due to the time difference of the receiver's internal clock and the GPS atomic timescale. With four GPS satellites, the three-dimensional position and the GPS time are determined through iterative computations. With additional satellites, the navigation solution is overdetermined, so least squares techniques are usually employed to obtain the best estimate for position. Various Kalman filter models obtain navigation solutions by smoothing through noisy measurements.

The military maintains a worldwide network of monitoring stations at Colorado Springs, Ascension Island, Diego Garcia, Kwajalein and Hawaii to collect GPS ranging data for all operational satellites. The Master Control Station uses the same signals collected by the GPS receiver and calculates the satellite ephemeredes and computes the future position of each GPS satellite and clock correction. The MCS calculates these prediction messages for at least 6 months and uploads these broadcast messages into each satellite on a daily basis. Each prediction message, which is valid for a few hours about the message epoch, does degrade with time, but the average root-mean-square (rms) satellite position error is about 2-3 meters after one day and approximately 24 meters after three days past an epoch.

1.2 GPS Ranging Signal

The GPS ranging signals are broadcast at 1575.42 MHz as L1 and at 1227.6 MHz as L2. The signals are broadcast synchronously, so that the receiver can directly calibrate the ionosphere group delay and apply the corrections to the measurements. More explanations concerning the deviations to the signals are given later in this thesis. Most civil GPS receivers use only the L1 signal, while the military GPS receivers are designed to use both L1 and L2.

The GPS signals can carry two modulations at the same time by phase quadrature. Right now, only L1 carries two modulations, while L2 carries one. The two modulations are:

(1) C/A or Coarse Acquisition Code. This is a short pseudorandom noise (PRN) code broadcast at a bit (or chipping) rate of 1.023 MHz on L1. This is the primary ranging signal, and it is always broadcast unencrypted. This code is repeated each millisecond, and each satellite broadcasts its own unique code allowing the receiver to track with a code tracking loop.

(2) P or Precise Code. This is a faster PRN code with a chipping rate of 10.23 MHz. Because P code has a higher modulation bandwidth, the code ranging signal is almost 10 times more precise. This reduces the noise in the received signal somewhat, but the signal offsets induced in the transmissions must still be eliminated. When the military encrypts P code, the new code is called Y Code, which is untrackable with the civil GPS receiver. This ensures the P code can not be spoofed by an adversary, and this feature is called antispoofing (AS). GPS receivers that can decrypt Y code are often designated as P(Y) code receivers.

In addition, the military's GPS design allows the degradation in the accuracy of the transmissions by perturbing the satellite clock or by incorporating errors in the broadcast satellite ephemeris. This degradation is called Selective Availability (S/A), and the formula for S/A is classified. However, toward the end of the Clinton Administration, S/A was set to zero for all GPS satellites, which is the continued situation we have now. This effectively is the situation prior to March 1990 when SA was first implemented for all GPS satellites.

There are two types of observations collected by most GPS receivers. One is the pseudorange, which equals the distance between the satellite and the receiver plus the equivalent distance offset due to the uncalibrated GPS clock internal in the receiver compared to the ultraprecise GPS timescale. Smaller corrective terms do create other offsets in the pseudorange, such as the ionosphere, the troposphere, multipath, Earth rotation, relativity, interchannel receiver biases, etc. The second observation type is the carrier phase, which is the difference between the received phase and the phase of the receiver oscillator at the epoch of measurement. Most GPS receivers make phase measurements at equally spaced intervals and keep track of the number of complete

cycles received since the beginning of the measurement. The accumulated phases observed since an arbitrary epoch is the typical output of this observation type.

The final component of the GPS signal is the 50 bit/second modulation to transmit information from the GPS satellite to the receiver. Each satellite transmits its location and the real-time offset to change the uncorrected GPS satellite clock value to the GPS timescale maintained at the MCS. Additional information is interleaved between the basic information, such as satellite health, locations of other GPS satellites, and the necessary information to lock onto P(Y) code after acquiring C/A code. Because of the slow transmission rate, it takes 12.5 minutes to get the complete information transmitted by any one GPS satellite.

Recent innovations in GPS receiver design in the civil community have allowed several techniques for effectively circumventing the effects of S/A so that C/A can approach the accuracy of P(Y) code receivers. However, most of these innovations suffer from a signal-to-noise (S/N) loss compared to direct P(Y) code tracking. Using these techniques, the civil GPS receivers generally operate in a low dynamic or nonjammed environment.

1.3 GPS Space Segment–Satellites

The original GPS satellites were called Block I and were deployed in 12 hour orbits inclined at 63 degrees in three planes that were spaced equally around the equator. The 63 degree inclinations took advantage of the natural stable equilibrium that maintained the argument of perigee in the orbits. The upgraded GPS satellites were the Block II and Block IIA models of the same external design, which were launched in 55 degree inclinations that the space shuttles could only deploy satellites due to fuel constraints. The number of orbital planes was increased to six, and the 24 operational satellites with fully functional spares now number around 30 (GPS JPO, 1998). The current GPS replacement satellites are Block IIR, and the future Block IIF will be deployed by 2008 (GPS JPO, 2006). The satellites use four electronically monitored momentum wheels for stabilization in three dimensions. The solar panels provide the basic power for the entire satellite operations. The ranging signal is radiated by a multiple element antenna that shapes the beam to broadcast the power in a 14 degree

cone toward the Earth with enhanced power concentrated toward the limb of the Earth. Wherever a GPS receiver may pick up radiation from a GPS satellite, this beam design allows nearly constant power at all local elevations on the Earth.

Figure 1: Block II or IIA Satellite

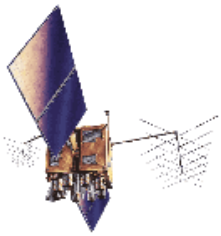

Figure 2: Block IIR or IIM Satellite

Figure 3: Block IIF Satellite

All GPS satellites carry multiple atomic clocks for redundancy since the cesium or hydrogen masers in the clocks eventually lose the gas of excited atoms required to generate the necessary frequency. The requirement for such a stable clock can be roughly determined as follows. The speed of light travels 1 foot in about 1 nanosecond. If the goal is to maintain GPS with an error of 5 feet in one day, then the precise clock must be stable to 5×10^{-9} s / 86400 s = 5.9×10^{-14} s/s, which only atomic clocks can provide. The atomic clocks with the frequency synthesizers synchronize the GPS signal generators and control the radio frequency centering of the two L band frequencies (1575.42 MHz and 1227.6 MHz). The atomic clocks synchronize the modulation of the navigation data onto the carrier signal for the broadcast of satellite time and position data

plus the other satellite status information. A more detailed explanation of the message content and signal transmission is given in a later section of this thesis.

1.4 GPS Control Segment–Monitor Stations

The United States Air Force maintains the control and monitoring of the GPS satellites after each has been initially deployed and initialized for space operation. This ground operation of GPS is called the GPS operational control segment. There are five automated globally distributed radio frequency stations deployed to control and continually monitor the GPS satellites. These ground monitor stations are located at Ascension Island, Diego Garcia, Guam, Hawaii and Colorado Springs, Colorado. The manned Master Control Station (MCS) is located in the Consolidated Space Operation Center (CSOC), which is about 15 miles east of Colorado Springs. The MCS maintains each of the satellites in its proper orbit with small maneuvers, which are infrequently required. The MCS makes corrections and adjustments to the satellite clocks as needed and can command major satellite maneuvers to relocate GPS satellites to fill in GPS coverage due to a gap after a satellite failure. The MCS continually collects ranging data from all monitor stations of the GPS satellites, computes new long-term satellite ephemeris and clock translation data, and uploads the navigation data daily covering 6 months into each Block IIR GPS satellite and 14 day uploads in older GPS satellites. These uploads can be accomplished at every monitor station except Hawaii, which only has the ground collection station.

The MCS also monitors the status and health of every GPS satellite. Part of the broadcast is the User Range Error (URE), which is the error projected to the user from each particular satellite. It should be noted that most of the URE is attributed to the satellite clock errors, and that any clock error is indistinguishable from range errors. The specified URE is computed as a weighted root-mean-square (rms) sum of radial, in-track and cross-track perturbations.

1.5 GPS User Segment–Receivers

A GPS receiver has several functions: antenna, preamplifier, reference oscillator, frequency synthesizer, downconverter, an intermediate frequency (IF) section, signal processing and applications processing. In most cases, the GPS receiver computes

navigation solutions and performs navigational tasks. Other applications employ time transfers, attitude determination, differential surveying or even simple GPS data collection only. So, the last functional area is more appropriately called applications processing to cover these other functions of the GPS receiver. A brief explanation of these alternate applications is warranted here.

Two users at previously known surveyed locations using two separate GPS receivers accomplish time transfers. The time transfer receivers simultaneously track a common satellite for an extended time interval (typically 13 minutes) and determine the time difference between an external ground clock input to the time transfer receivers and the GPS clock. By exchanging time differences between each ground receiver and the common GPS satellite clock, a second subtraction between the difference data shows the time difference between the ground clocks since the common GPS satellite clock cancels out. This is called a time transfer, which mimics the actual transport of an ultraprecise clock to another clock's location for comparing the actual clock differences needed in calibration and timekeeping.

Attitude determination uses multiple GPS antennas with one GPS receiver that computes the length difference in the ranging data between all pairs of antennas. The GPS receiver computes the ratio of the measured line-of-sight projection from the received time differences against the actual distance between each pair of antennas to get the geometric three-dimensional angles of the antenna platform relative to the GPS satellites. A direct angular rotation in cosines then obtains the orientation relative to local level coordinates.

The differential surveying techniques resort to placing a GPS receiver at a known benchmark as the base station and computing the expected GPS ranges between each visible GPS satellite and the benchmark against the measured GPS pseudoranges. These differences are treated as corrections to the GPS pseudorange broadcasts and are either transmitted to any nearby remote GPS receiver or incorporated in postprocessing of the remote receiver's GPS data. This practice is called differential GPS (DGPS), and it allows the surveying community to obtain centimeter accuracy in GPS surveys over long baselines relative to a benchmark.

However, the vast majority of GPS receivers are used for navigation, and this processing is discussed later in detail. A special feature of the GPS receiver is the ability to create an extremely precise ranging signal by reproducing and tracking the radio frequency (RF) carrier. Because the wavelength of L1 is 19 cm, tracking it to $1/100^{th}$ of the wavelength would provide a precision of 2 mm. Unfortunately, precision is not accuracy. To provide the same accuracy as the precision, the GPS receiver must determine which whole cycle or revolution in the phase has occurred since the signal was transmitted from the GPS satellite, which is called integer cycle ambiguity resolution. The surveying community uses multiple stationary receivers and processes the measured ranging data with double or triple differencing to resolve this ambiguity. Strong GPS signals above 40 dB-Hz are desirable for this processing to measure precisely the fractional part of the cycle phase. Dynamic users have a much harder problem resolving cycle ambiguity with far more cycle slips coming from multipath, ionosphere distortions and other error sources. What will be attempted in this thesis is to use a single GPS receiver to obtain similar accuracy in the ranging data by a novel process.

1.5.1 Functions of GPS Receiver Architecture Components

The antenna may have a single or multiple element components with the necessary control electronics. Depending on the performance requirements, it may be passive or active, although most GPS antennas are passive. The antenna's function is to receive the GPS signals while, if appropriately designed, reject multipath and interference signals.

The preamplifier generally contains burnout protection, filtering, and a low-noise amplifier (LNA). Its primary function is to set the receiver's noise figure and reject out-of-band interference.

The reference oscillator provides the time and frequency reference for the receiver. All GPS receiver measurements are based on the time-of-arrival of the pseudorandom noise (PRN) code phase and received carrier phase and frequency information. So, the reference oscillator is the key component of the receiver. It regulates the frequency synthesizer and drives the tracking loops internal to the receiver. Often, the same oscillator is used by the downconverter to convert the radio frequency

(RF) inputs to intermediate frequency (IF) output, which is easier to process and measure than direct RF signals.

The IF electronics reject interference when processing out-of-band noise. It also increases the amplitude of the signal and reduces noise to a workable signal level. The IF electronics may include automatic gain control (AGC) circuits to maintain an adequate signal range and to suppress pulse-type interference.

The signal processing function: splits the incoming signal into multiple channels simultaneously, generates the reference PRN codes of the signals, acquires the satellite signals, tracks the code and carrier of the satellite transmissions, demodulates the data from the signals, determines the pseudorange measurements from the PRN code, extracts the carrier frequency and carrier-phase measurements, obtains the signal-to-noise ratio (SNR), and estimates the GPS system time. The output of the signal processing are pseudoranges, pseudorange rates from the carrier frequencies, and delta pseudorange measurements from the carrier phase values, SNR data, internal receiver time tags, and GPS system information of all visible satellites.

The applications processing uses its output to satisfy its design and application requirements. Some of the diverse GPS applications include time and frequency transfers, static and kinematic surveying, ionospheric total electron content (TEC) monitoring, differential GPS reference station receivers, and GPS satellite integrity monitoring for detecting and alarming during anomalous satellite behavior. Still, the GPS is primarily a satellite navigation system, and most GPS receivers are navigational.

1.5.2 GPS Signal Offsets: Perturbations in GPS Ranging

The GPS signal frequencies L1 and L2 were chosen within the frequency band to minimize the absorption of the signal power through the atmosphere. This choice allows the use of omnidirectional antennas without requiring pointing to the satellites. Signal attenuation from rainfall or fog does not interfere appreciably with acquisition and tracking. However, the atmosphere does cause enough deviation that compensations are required for this level of accuracy. The total amount of deviation is a function of the intensity of the propagation effects, the local elevation angle that the signal path travels through the atmosphere, and location of the GPS receiver. The general order of

magnitude due to these offsets is ionospheric group delay, troposphere group delay caused by wet and dry atmospheres, relativity effects from special and general relativity, and multipath, which are discussed separately with more detail.

The ionosphere contains ionized gases that vary widely from day and night times and with solar radiation fluxes. The ion plasma is mostly electrons, and the total electron content (TEC) affects both the group and phase velocity of the GPS signals. The peak ionosphere electron content is somewhere between 200 and 400 km thick and can vary by 100 times between day and night (Klobuchar, 1996). The refractive index of the ionosphere is a function of the frequency, but the refractive index for the troposphere affects both L1 and L2 equally. To first order, the ionospheric delay varies inversely with frequency squared. The effect can vary from 2 to 50 ns of group delay for a satellite at zenith. At lower elevations, the effect can be 3 to 4 times worse excluding periods of maximum solar sunspot activity.

The troposphere is a region of dry gases and water vapor that extends up to 50 km above the Earth. The index of refraction is slightly greater than unity and nearly equal to free space. This region is not generally ionized to any appreciable effect. The excess group delay is typically about 2.6 meters for a satellite at zenith, but at low elevations below 15 degrees, the troposphere can increase the offset to 20 m.

Relativity effects have been calculated and demonstrated in uncorrected GPS signal transmissions. The mean gravitational red shift and the mean velocity of the satellites has been removed by a slight offset at the factory in the actual satellite transmission frequency, so that the received signal at Earth's surface is the ideal L1 and L2 frequencies after the satellite has been deployed. Because the satellite orbit is always an ellipse and not ideally circular, the gravitational potential and satellite velocity vary predictability, and a formula has been derived to eliminate the remaining relativistic effects in the GPS ranging information. If no relativistic compensation was implemented, then the offset in frequency or equivalent clock rate would create ranging errors over 28 km in 24 hours after the last message upload to the satellite (Ashby, 1990).

Finally, multipath is highly dependent on the reflective environment that can interfere or smear the GPS signal. Generally, the direct signal is stronger than any

reflected signal from multipath, and the GPS signal processing usually eliminates such unintentional interference, but exceptions do exist. For example, the reflected signal's amplitude from the sea surface can be at times as large or larger than the direct signal. When the multipath reflects a delayed signal that appears within the code chip, this superposition tends to skew the resolution of the measured pseudorange signal (a type of smearing) and degrade the accuracy somewhat in the navigation solution.

Chapter 2 GPS Signal Structure and Performance

The GPS signal is a type of spread spectrum signaling, which the fundamentals are discussed. A detailed explanation of the coarse/acquisition and precision codes is given. Also, the signal to noise ratio and the signal performance characteristics are discussed with performance bounds on C/A and P(Y) code pseudorange tracking.

2.1 Spread Spectrum Signal Fundamentals

Spread spectrum signaling is a method to add a data signal of bandwidth B_d by modulating it onto a sinusoidal carrier transmission and then spreading its bandwidth to a much larger value $B > B_d$. The bandwidth spreading can be accomplished by multiplying the data-modulated carrier by a wide bandwidth-waveform s(t) for spreading. A data bit stream, D(t), of 1 and –1 values and a clock rate f_d are first modulated on a carrier signal of power P_d to form the narrow bandwidth signal:

$$d(t) = D(t)\sqrt{2P}\cos\omega t$$

This signal of bandwidth B_d is then spread by a binary pseudorandom signal s(t) where $s(t) = \pm 1$ and clock rate f_c that greatly exceeds the data bit rate ($f_c >> f_d$). The data and spreading waveforms have the power spectral density of:

$$G(F) = \frac{1}{f}\left[\frac{(\sin \pi F/f)}{\pi F/f}\right]^2$$

Because the time of the data and clock transition are synchronous, the spread spectrum product D(t)s(t) has exactly the same spectrum as that of s(t) only with the following form.

$$s(t)d(t) = s(t)D(t)\sqrt{2P_d}\cos\omega t$$

The spreading signal is $s(t) = \sum_{n=-\infty}^{\infty} S_n p(t - nT_c)$ where p(t) is a rectangular unit pulse over the interval $[0,T_c]$ as $1/T_c = f_c$ for the frequency of the clock rate and S_n is a pseudorandom code sequence. Generally, different spreading waveforms can separately modulate in-phase and quadrature carrier components. For GPS, the discussion is restricted to rectangular pulses and biphase modulation. This form of spread spectrum is

called direct sequence-spread spectrum, which is one of the different forms of spread spectrum. This form of spread spectrum provides a means to recover precise timing and the pure RF carrier.

The direct sequence-spread spectrum signal then passes through a channel with additive white noise n(t) of power spectral density N and interference b(t) to form the receive signal $r(t) = s(t)\, d(t) + n(t) + b(t)$ where b(t) is a pure tone interference of power P_b. Inside the receiver, an identical replica of the spreading signal s(t) is generated and correlated by mixing and filtering with the received noisy signal. The signal has a narrow autocorrelation envelope of width inversely proportional to the clock rate f_c. The multiplier converts the spread signal into $d(t)\, s^2(t) = d(t)$ since $s^2(t) = 1$. The multiplier compresses the spread spectrum signal back to its original narrow bandwidth leaving only the data modulation. The narrowband interference b(t) has been spread to look like s(t) due to the multiplier, similar to the manner which the narrowband signal d(t) was spread in the transmitter. Filtering this multiplier output through a bandpass filter passes the narrowband signal d(t) relatively undistorted with only a fraction of the noise and interference power following d(t) (Spilker, 1996).

Demodulation of this filter output produces a bit error rate that is determined by this noise and interference level. If only thermal noise was present, then the effects cancel the spreading and despreading by the pseudorandom (PN) code s(t) in the transmitter and receiver. So, the use of properly synchronized spread spectrum signaling neither improves nor degrades the signal performance against a thermal noise background.

But, the performance against a steady interference is greatly improved, because the interference power level is reduced by the ratio of the clock rates f_c/f_d, where f_c is the PN chip rate regulated by the internal clock and f_d is the data bit rate. The ratio, f_c/f_d, is termed the processing gain of the spread spectrum system, because that ratio determines what fraction of the interference power passes through the output. This property of spread spectrum signaling is effective against even broader classes of interference.

2.2 Autocorrelation Functions and Spectrum

A random sequence of equally probable events, each lasting the time interval $T_c = 1/f_c$, the inverse of the clock rate, has a triangular autocorrelation function and a power spectral density shaped as $(\sin(\pi\tau f)/\pi\tau f)^2$. Pseudorandom codes or sequences can be generated with suitable feedback shift registers. For GPS, $T_c = 20$ ms for the chip length.

2.3 Multiple Access with Spread Spectrum Signaling

The GPS design takes advantage of using multiple signals simultaneously to get access on the same frequency channel with minimal interference between signals. Spread spectrum signaling has the capability of providing code division multiple access (CDMA), provided the number of signals M is not too large. Even though all GPS satellites are not equally distant from the GPS receiver, a judicious choice in pseudorandom codes can be nearly uncorrelated for all possible time offsets in the various multiple M transmissions from M satellites. For example, if all M signals are received with exactly the same code clock delay, it is possible that multiple access interference is negligible provided $M \leq f_c T_d$ where $f_d = 1/T_d$ is the clock bit rate. An analysis of this feature is described next.

Suppose two multiple access signals $s_1(t)$ and $s_2(t)$ have uncorrelated pseudorandom codes and are transmitted on the same frequency channel and received with independent random timing. The receiver mixes the input with the desired reference signal code $s_1(t)$ to get a correlator output of $s_1(t - \tau_1)\,[s_1(t - \tau_1) + s_2(t - \tau_2)] = 1 + s_1(t - \tau_1)\, s_2(t - \tau_2)$. The unit term is the objective, and the multiple access term is $s_1(t - \tau_1)\, s_2(t - \tau_2)$, which can be defined as $G_{ma}(f)$.

$$G_{ms}(f) = P_s \int G_s(\upsilon) G_s(\upsilon - f)\, d\upsilon \text{ and}$$

$$G_{ma}(0) = P_s \int G_s^2(\upsilon)\, d\upsilon$$

Assume the processing gain is large where $f_c/f_d >> 1$. Then, only the multiple access interference spectrum near $f = 0$ is significant, because the correlation filters have a bandwidth on the order of f_d. The integral can be computed.

$$G_{ma}(0) = P_s \int_0^{\infty} \left(\frac{\sin(\pi f)/f_c}{\pi f / f_c} \right)^4 df = \frac{2}{3}\frac{P_s}{f_c}$$

If the signal is filtered to include only the main lobe, the factor increases from 2/3 to about 0.815. If the signal spectrum is rectangular, the factor is unity. The quantity E_b is the energy per bit $E_b = P_s/f_d$. The multiple access equivalent noise density can be calculated to determine how many M-1 interfering access signals can be tolerated in the presence of the desired signal for each tracking channel.

$$N_{oeq} = N_o\left[1 + \frac{(M-1)P_s}{N_o}\right] \int_{-\infty}^{\infty} G_s^2(f)df \qquad \text{for } G_s(f), -\infty < f_o < \infty$$

$$= N_o\left[1 + \frac{(M-1)P_s}{f_o N_o}\right] \qquad \text{for rectangular spectrum, } \frac{P_s}{2f_o}, -f_o < f < f_o$$

For biphase modulated signaling, the E_b/N_{oeq} term should be 10 or greater if no error correction coding is employed, as is done in the open GPS design. Then, the number of equal power multiple access signals is limited by

$$M < \frac{N_o f_c}{P_s} + 1 = \frac{N_o f_c}{E_s f_d} + 1 = \frac{1}{10}\left(\frac{f_c}{f_d}\right) \quad \text{for} \quad \frac{E_s}{N_{oeq}} = 10$$

For example, if $f_c = 10^6$ Hz and f_d = 50 bit/second, then M < 2001. This is one desirable characteristic in the spread spectrum transmission that allows GPS receivers to pick out the desired GPS signal from multiple transmissions of different codes over the same carrier.

Depending on the choice of the number of registers used to generate the code, the number of states available is $2^m - 1 = n$ for m binary shift registers. The autocorrelation of the pseudorandom sequence where s(t) = ±1 is:

$$R(i) = \frac{1}{n}\int_0^n s(t)s(t+i)\,dt$$

When the maximal length sequence becomes more random in appearance, then the spectrum approximates the $(\sin z/z)^2$ of the random sequence, and the autocorrelation function is a sharp, narrow peak. This autocorrelation waveform with a high clock rate

signal can be used for very accurate measurements of time and range, which is characteristic of GPS.

As mentioned before, GPS satellites transmit C/A and P(Y) codes continually. Time multiplexing the two codes was not adopted (i.e., transmitting the short C/A code completely then switch to a portion of the long P(Y) signal). This option would have disrupted the continuous carrier phase measurements, and time gating would have lowered the cross-correlation performance of the GPS transmissions. The GPS signal design had the civil C/A signal modulated on the in-phase component of the L1 carrier and the P(Y) code of the military modulated on the quadrature phase (90 degree rotation). This allows a nearly constant envelope around the carrier even though the signal strength of C/A is 3 dB higher than P(Y). The GPS signal without the modulation has the carrier form $XP_i(t) \cos(\omega_o t) + XC_o(t) \sin(\omega_o t)$.where XP represents the magnitude of the P(Y) code and XC represents the C/A code magnitude. All data are biphase modulated identically on both in-phase and quadrature components.

2.4 GPS Radio Frequency Selection and Signal Characteristics

The L-band was chosen for several reasons. Both the satellite transmission powers and received signal powers were acceptable, and the satellite antenna patterns covered the Earth quite effectively. C-band paths had nearly 10 dB higher losses when using the omnidirectional receiver antenna with the fixed transmit antenna beamwidth and range. The ionospheric delay and fluctuation affected UHF much worse, and there was too much use in UHF to obtain two large 20 MHz bandwidth frequency assignments from the international radio regulatory commission.

Each of the L1 and L2 center frequencies is a coherently selected multiple of 10.23 MHz master clock. All the signal clock rates for the codes, radio frequency carriers and the 50 bps navigation data stream are coherently related. In particular,

$$L1 = 1575.42\text{MHz} = 154 \times 10.23\text{MHz}$$
$$L2 = 1227.60\text{MHz} = 120 \times 10.23\text{MHz}$$

The ratio of L1/L2 = 77/60, which is exploited in the enhancement design proposed in this thesis. The ionosphere delays and corrections are also discussed in a separate

section. The relativity effects in the GPS transmissions are explained in a later section as well.

2.4.1 GPS P and C/A Codes

The L1 signal is modulated by both a 10.23 MHz clock rate for the P(Y) code signal and by a 1.023 MHz C/A code. The binary modulating signals are formed by adding the 50 bps binary data D to form C/A⊕D on the in-phase carrier and P⊕D on the quadrature. The peak power of the C/A signal exceeds the peak power of P(Y) code by 13 dB, because C/A is 3 dB stronger and has 1/10 the chip rate and bandwidth of P(Y) code. The rms clock transition time difference between the C/A and P(Y) code clocks is required to be less than 5 ns. Both C/A and P codes are from a class called product codes where each is the product of two different code generators clocked at the same rate.

The P code for any GPS satellite is the product of 2 pseudorandom codes X1(t) and X2(t+nT), where X1 has a period of .5 seconds or 15,345,000 chips, and X2 has a period of 15,345,037 or 37 chips longer. Each of these numbers is relatively prime, so the product is $2.35469592765 \times 10^{14}$, which is slightly more than 38 weeks long. For GPS, this overall period has been subdivided so that 37 possible GPS transmitters gets a one-week interval of code, which does not overlap with any other GPS transmitter.

The C/A code is relatively short with a period of 1023 bits or one ms duration at the 1.023 Mbps bit rate. This short code permits rapid acquisition with only 1023 bits to search. Each C/A code is the product of two 1023 bit pseudorandom codes G1(t) and G2(t). Only 37 codes are defined for GPS, so as to maintain uniformly low cross-correlation with the sidelobes for all other needed satellite codes and possible offsets. Epochs of the G code occur at 1 kbps and are divided down by 20 to get the 50 bps data clock. All clocks are in phase synchronism with the X1 clock of the P code.

2.4.2 GPS Data Format

Using the 50 bps data bit rate, there are five subframes of six seconds each for a total frame period of 30 seconds. Each subframe has a Telemetry (TLM) word at the beginning for synchronization identification. A Handover Word (HOW) that contains the Z-count follows the TLM word for handover from the C/A code to the P code. The

extraordinarily long P(Y) code makes it extremely difficult to acquire without assistance, because a receiver correlator must be timed to within one P-code chip or 0.1 μs for synchronizing. The number of the X1 epochs from the P(Y) code is the Z-count since the beginning of the week. There are four X1 epochs per data subframe of 6 seconds. When multiplied by 4, the HOW equals the Z-count at the beginning of the next 6-second subfame. With timing from the relatively short C/A code and knowing the subframe epoch time and HOW words, the P-code at the next subframe epoch can be readily acquired.

The 5 subframes contain the data transmitted from the GPS satellites, which takes 12.5 minutes to complete the entire transmission. Subframe 1 contains the satellite clock corrections and satellite quality, Subframes 2 and 3 list the satellite's ephemeris, and Subframe 4 contains, among other things, the satellite's almanac with ionosphere and UTC corrections. Subframe 5 has 25 different messages with each message contained in a separate 30 second frame. So, each GPS satellite transmits a 30-second record or frame of data that contains the first four subframes of information about itself and sends the fifth subframe about one of the other satellites. Thus, each satellite eventually broadcasts the entire almanac for all the other GPS satellites in 12.5 minutes. The almanac is very useful for predicting weeks and months into the future GPS when predicting satellite coverage above the local horizon so that appropriate steps can be taken with the receiver tracking loops to compensate Doppler shifts in the GPS signals.

2.4.3 Estimate of Delay Measurement Noise on Code and Carrier

Estimate measurement error by white noise added to the carrier over the bandwidth:

$$r(t) = \sqrt{2P_s}\sin(\omega_o t + \phi) + N_s(t)\sin(\omega_o t + \hat{\phi}) + N_c\cos(\omega_o t + \hat{\phi})$$

Multiply it with a nearly pure quadrature coherent reference phase signal (e.g. $2\cos(\omega_o t + \hat{\phi})$ and the phase detection through a low-pass filter to obtain:

$$r(t)2\cos(\omega_o t + \hat{\phi}) \approx \sqrt{2P_s}\sin(\phi - \hat{\phi}) + N_c t \approx \sqrt{2P_s}(\phi - \hat{\phi}) + N_c \text{ for } \left|\phi - \hat{\phi}\right| << 1.$$

The one-sided noise power spectral density of $N_c(t)$ is $2N_o$. If the phase error $\phi - \hat{\phi}$ is filtered to a bandwidth of B_L, then the output phase difference plus phase noise can be written in a normalized form $\phi - \hat{\phi} + \phi_n$, where

$$\phi_n^2 = \frac{N_o B_L}{P_s} \quad \text{or as an equivalent rms time error} \quad \sigma_T = \frac{1}{2\pi f_o}\frac{1}{\sqrt{S/No}}.$$

S/No is the equivalent signal-to-noise ratio in dB. The GPS carrier phase noise is specified to be sufficiently low to allow tracking with a 10-Hz noise bandwidth. If a carrier-tracking loop of bandwidth B_L=10 Hz is employed and the C/No is 45.2 dB, then a signal-to-noise ratio in the loop under linear conditions is S/No = C/No B_L = 45.2 – 10 = 35.2 dB, since at frequency f_o, $\sigma_\phi = 2\pi f_o \sigma_T$. The S/No corresponds to a rms phase noise σ_ϕ where:

$$\sigma_\phi^2 = (\sigma_T 2\pi f_o)^2 \approx \frac{1}{S/No} \quad \text{or} \quad c\sigma_T = \frac{\lambda}{2\pi}\frac{1}{\sqrt{S/No}}$$

when the S/No for the loop is sufficiently high. If S/No = 35.2 dB or 3311 and the wavelength λ = 19.05 cm for L1, then the output phase error in distance terms is $c\sigma_T \approx 0.5\text{mm}$.

So, the carrier-tracking phase noise measurement error from thermal noise is about 0.5 mm. Other carrier phase noise and dynamic tracking errors would increase this estimate to several mm, but the carrier phase measurement error is still very small. It had been considered that carrier phase measurements had an ambiguity in absolute number f phase cycles and could only be included as differential phase measurements. If no cycles are slipped or miscounted, the period of satellite visibility can continue for several hours of tracking. The accumulated delta range (ADR) can be added continuously from carrier phase measurements over time and converted into distance by multiplying by the speed of light. The ADR curves are smooth with a bias offset from the ionosphere, but they have the same shape as the pseudorange code measurement curves, except the code curve has much higher noise.

Code tracking noise is on the similar order of $f_c = 1/T_c$, = 1.023 MHz for the C/A code. If the received code-tracking signal is converted to baseband and multiplied by a delay replica of the signal, then the output code signal is:

$$\sqrt{2P_s}s(t+\tau)s(t+T_c/2+\hat{\tau})+N_{cl}(t)\approx\sqrt{2P_s}\left(\frac{\tau-\hat{\tau}}{T_c}\right)+N_{cl}(t)+\text{constant}$$

by replacing the sine term with the small angle $|\tau-\hat{\tau}|<<T_c$. A crude estimate of the rms delay error is then

$$\sigma_T=\frac{T_c}{\sqrt{S/No}}=\frac{T_c}{\sqrt{P_s/N_oB_L}}$$

With a code-tracking loop bandwidth of 3 Hz, C/No = 45.2 dB-Hz and all terms are linear, then the code tracking loop is S/No = 45.2 – 4.8 = 40.4 dB. For this nonoptimum code-tracking measurement, the rms tracking error is:

$$\sigma_\tau\approx T_c\frac{1}{\sqrt{S/No}}=9.32\text{ ns}\quad\text{or}\quad c\sigma_\tau\approx 2.8\text{ meters}$$

when S/No >>1 and T_c = 0.9775 μs for the 1.023 Mbs C/A code. Code-tracking errors are larger than carrier-tracking errors by roughly by the ratio of carrier frequency to code clock frequency. These estimates are only intended as crude estimates to give a rough idea of the true accuracy when collecting measurements.

Chapter 3 GPS Receiver Equipment Architecture

Generally, the GPS receiver equipment does two measurement operations. The first is the receiver tracks the GPS signals, which, in this application, is accomplished with a delay lock loop (DLL) for measuring both pseudorange. The second is the receiver measures the accumulated delta range (ADR), which is a pseudorange-rate measurement using the carrier. Most GPS receivers have one antenna with a low-noise amplifier, although multiple antennas can be used either for GPS attitude determination receivers or for beam steering to enhance reception and exclude potential jamming sources. The antenna output is fed into a combined radio frequency bandpass filter and low-noise amplifier for amplifying the signal and filtering out high-level interfering signals from nearby frequencies. The radio frequency filters must have low signals losses and maintain sufficient bandwidth and phase linearity to minimize distortion of the C/A and P(Y) code signals. The radio signals pass through many stages of amplification, downconversion, intermediate frequency (IF) amplification and filtering, analog to digital (A/D) measurements, and finally sampling with quantization. This last stage can be done either in the IF or baseband regions. The in-phase (I) and quadrature (Q) samples are taken from the input signals with noise. All of these operations can be combined in a single monolithic microwave integrated circuit chip (MMIC).

The I and Q samples are fed to a bank of DLLs, which each channel tracks a different satellite, measures pseudorange and recovers the carrier which is biphase modulated with the GPS broadcast data. Each channel with its DLL obtains values for the pseudorange, carrier phase and navigation data using digital processing. Some advanced GPS receivers have 12 or more parallel DLLs channels to track nearly every possible visible satellite at one time. The receiver collects the parallel pseudoranges, the carrier phase measurements, and the navigation data where each satellite position is calculated and the pseudorange and clock corrections are made. As discussed later, the receiver corrects for ionosphere delay, troposphere delay, relativistic effect, Earth rotation effects and equipment delays. All of these data are then processed by an Extended Kalman filter to smooth the data results and compute a solution at the antenna's phase center. If other components are part of the total navigation system, such as an

inertial measurement unit (IMU), then a lever-arm translation of the GPS navigation solution is usually made to combine the GPS result with other navigation subsystems.

3.1 Coherent Early/Late Delay Lock Tracking

Both coherent code and carrier tracking are performed. The received signal has additive white noise. The signals should be bandlimited with finite rise time signals and with a continuous autocorrelation function that is effectively differentiable. The delay lock loop uses an identically generated signal that is time offset, which determines the total delay between the GPS satellite clock at transmission to the receiver's internal clock at reception. These operations are performed in discrete time digital operations. The timing is accomplished by the finite rise time of the correlated waveform as it crosses the reference power level. This demodulated waveform resembles a trapezoidal function with rounded corners instead of the ideal step function. The correlated signal makes its transition between levels (i.e. transitions from –1 to +1 or +1 to –1). The only portion of the waveform useful for time delay measurement is the transient rise time, measured usually at the half power point. This DLL is a suboptimal design compared to nearly all other communication systems using all of the waveform. The portion of the overlapped code maintains power when mixing the input and generated codes, but the unmatched portion will cancel out between the input and generated codes due to the random nature of the code. The autocorrelation function is the result of slewing the generated code.

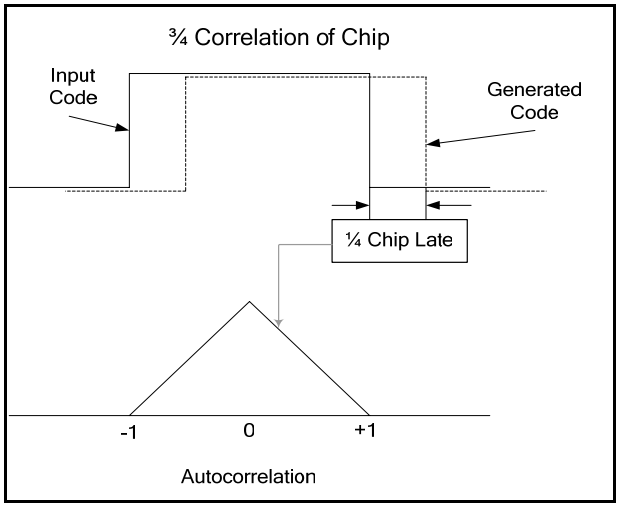

Figure 4: Correlation Curve between Input and Generated Code Tracking Loop

The simple solution to this deficiency with the DLL is to incorporate two delay lock loops, one that has its reference waveform set early by T/2 of a chip width T and the other set late by T/2 of the chip width. Here, a chip refers to the code length of either the C/A code or P(Y) code between transitions. The output two signals are differenced to produce the half-early minus half-late DLL pattern shown in Figure 5. The sharp slope gives a much finer resolution at the zero crossover point, which is the reception time of the incoming signal.

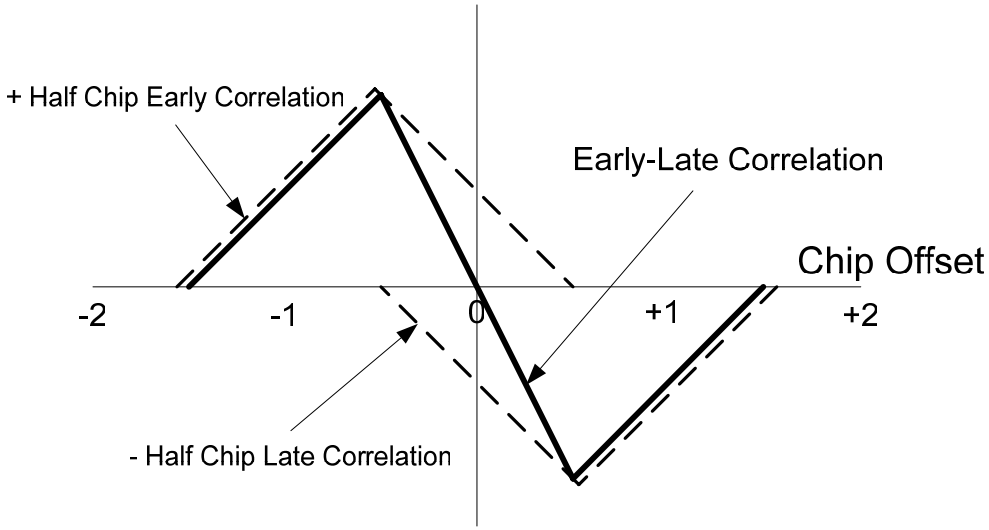

Figure 5: Early – Late Correlation Curve for Half Chip Spacing

Chapter 4 GPS Satellite Navigation Broadcasts

From first principles using Newton's second dynamical law ($\vec{F} = m\vec{a}$) and Newton's law of gravitation $\left(F = k\,mM/r^2\right)$, one can derive the equations of motion for a satellite orbiting a uniform spherical Earth as $\ddot{\mathbf{r}} = -kM\mathbf{r}/r^3$ where the acceleration $\ddot{\mathbf{r}}$ is collinear with the position vector $\mathbf{r}$. No acceleration exists perpendicular to the position vector with no velocity in that direction either. Convert the two dimensional vector between Cartesian x and y coordinates to polar coordinates of r and θ to obtain two differential equations of motion. One is immediately integrated so that $r^2\dot{\theta} = \text{constant}$, which shows that angular momentum is conserved and orbital motion is constrained to planar motion. The steps to integrate the other differential equation for this two-body problem are documented in many intermediate physics textbooks (Marion, 1970). The result is the elliptical equation for radial motion, $r = \dfrac{a(1-e^2)}{1+e\cos\upsilon}$ where a is the semimajor axis, e is the eccentricity and υ is the true anomaly angle. Other variables that are more convenient for calculations are eccentric anomaly E and the mean anomaly M where $E - e\sin E = M$ and $M = n(t - t_0)$ with n being mean motion such that $n = \sqrt{\dfrac{\mu}{a^3}}$ and μ is the reduced mass. The relationship between true anomaly υ and E is $\tan\upsilon = \sqrt{1-e^2}\,\sin E/(\cos E - e)$. The next figures (Jeyapalan, 1993) illustrate the orbital geometry of these six parameters (semimajor axis a, eccentricity e, argument of perigee ω, inclination i, right ascension Ω, .and mean anomaly M):

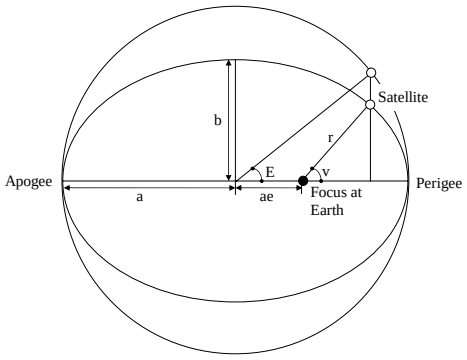

Figure 6: Orbital Plane Parameters

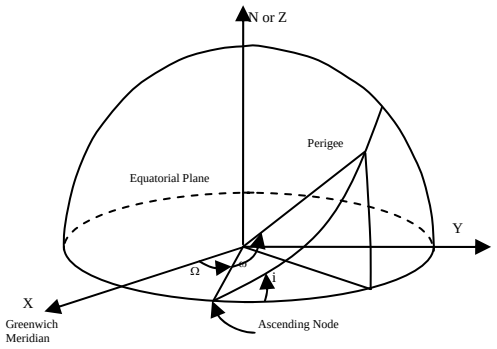

Figure 7: Parameters Orienting Orbital Plane

However, the world is neither homogeneous in its density nor perfectly spherical in shape. Letting V be the gravitational potential, the problem of solving for the orbital equations of motion can be obtained from Laplace's equation: $\nabla^2 V = 0$. In spherical coordinates, the partials can be separated, which the most difficult is the Legendre equation with a series solution containing constants called Legendre or associated Legendre polynomials. The complete real solution (Kaula, 1966) of the Laplace equation is:

$$V = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \frac{1}{r^{l+1}} P_{lm}(\sin\phi)[C_{lm} \cos m\lambda + S_{lm} \sin m\lambda]$$

The spherical harmonics are the products of the Legendre polynomials with each trigonometric expression in the square brackets. The spherical harmonics are mutually orthogonal, so that the general representation of a gravitation surface potential is a series of spherical harmonics, analogous to the Fourier series for a function in rectilinear space. It is noted here that the C_{20} term is a thousand times larger than any subsequent coefficient in the series.

Fortunately, the noncentral gravitational field of the Earth is slightly perturbed from the pure central force problem. So, the Keplerian ellipse and its orientation can be used as a coordinate system with 6 parameters of the Keplerian ellipse replacing the six terms of position and velocity of the earlier central force problem. Laplace's equation is changed from 3 second order differential equations into 6 first order partial differential equations. The method of solution is through the use of Lagrangian brackets to set up 6 simultaneous solutions of the Laplace equations. The derivation (Kaula, 1966) is very

involved with the 6 Keplerian elements exhibiting a first derivative in terms of the forcing function F, which is not a central force, but a perturbing force. If the series solution is shortened to just the primary harmonic, $V_{2010} = \frac{\mu C_{20}}{a}\left(\frac{a_e}{a}\right)^2 F_{202}(i) G_{210}(e)$, the forcing function $F = \frac{2\mu}{a} + V_{2010}$ yields three nonzero first derivatives in ω, Ω and M. GPS orbits took advantage of the first derivative of the argument of perigee, ω, which is:

$$\frac{d\omega}{dt} = \frac{3 n C_{20} a_e^2}{4(1-e^2)^2 a^2}\left[1 - 5\cos^2 i\right].$$

The original GPS constellation was chosen in inclination to make this derivative be zero, which is why the Block I GPS satellites were in orbits of 63.435 degree inclinations. The mean anomaly deviation is an along-track error from the standard orbital prediction. Thus, the earliest GPS constellation had only to correct for the drift in right ascension with minimal orbital corrections. The first derivative in right ascension is given:

$$\frac{d\Omega}{dt} = \frac{3 n C_{20} a_e^2}{2(1-e^2)^2 a^2}\cos i$$

Subsequent satellites after the initial Block 1 satellites were placed by the Space Shuttle, which is limited to a maximum of 55 degree inclined orbits due to fuel limitation, which all subsequent and new operating GPS satellites are or will be placed.

Chapter 5 GPS Navigation Algorithms

The overall objective in GPS navigation is to determine the position in X, Y, and Z Earth-Centered, Earth-Fixed (ECEF) coordinates and time in the GPS timescale. These crucial terms and any other parameters are collectively the "state" of the system. In moving applications, the three components for velocity and three for acceleration are added. In some systems, other components such as an Inertial Measurement Unit (IMU) may be included. For this research, only the sole GPS receiver is being considered. The overall processing involves both the measurement model and a dynamic model that are discussed in detail. The current measurement model is constructed from a linearized form of the irrational equations that describe the geometry of the problem, and this formulization then relies on a converged iteration for solving the initial parameters. The dynamic model continues in the form of a linearized Kalman filter to predict the current dynamics of the system, which is used to predict the Doppler shifts in the GPS signals so that the tracking loops can follow the signal frequency without loss of the signal. Some problems may occur when the iterative technique converges prematurely with this standard method. To correct these problems, a new derivation of an exact direct navigation solution without iteration is given to replace the iterative technique.

5.1 Standard GPS Measurement Model

Ideally, the true distance between each GPS satellite and the receiver's antenna is computed by taking the time difference between the reception and transmission and multiplying that by the speed of light (c). This would be possible if there were no propagation delays caused by the ionosphere and troposphere and by relativity. Worst of all, the internal receiver's clock or oscillator is not calibrated to the GPS atomic timescale. Fortunately, this large timescale offset, when multiplied by the speed of light, can be solved with the position parameters. The clock offset with the actual range determines the measurement called pseudorange (PR). Let the x_s,y_s,z_s be the position of the satellite at transmission, x_r,y_r,z_r be the position of the receiver at reception and the time difference $b = (\tau_{rec} - \tau_{sat})$, and error ε, then the pseudorange can be expressed as:

$$\rho = \sqrt{(x_s - x_r)^2 + (y_s - y_r)^2 + (z_s - z_r)^2} + c(\tau_r - \tau_s) + \varepsilon_s = \sqrt{(\vec{s} - \vec{r})^2} + c(\tau_r - \tau_s) + \varepsilon_s$$

At least 4 satellite equations of this form are solved simultaneously to get the three dimensional position and time offset. The standard method linearizes this system of equations about some nominal value for the best estimate. By taking a Taylor series expansion about the nominal value, only the first order term is kept and the result is:

$$\Delta\rho = \hat{\rho} - \rho_o = \begin{bmatrix} -\frac{(x_{s1}-x_r)}{\sqrt{(\vec{s}_1-\vec{r})^2}} & -\frac{(y_{s1}-y_r)}{\sqrt{(\vec{s}_1-\vec{r})^2}} & -\frac{(z_{s1}-z_r)}{\sqrt{(\vec{s}_1-\vec{r})^2}} & 1 \\ -\frac{(x_{s2}-x_r)}{\sqrt{(\vec{s}_2-\vec{r})^2}} & -\frac{(y_{s2}-y_r)}{\sqrt{(\vec{s}_2-\vec{r})^2}} & -\frac{(z_{s2}-z_r)}{\sqrt{(\vec{s}_2-\vec{r})^2}} & 1 \\ -\frac{(x_{s3}-x_r)}{\sqrt{(\vec{s}_3-\vec{r})^2}} & -\frac{(y_{s3}-y_r)}{\sqrt{(\vec{s}_3-\vec{r})^2}} & -\frac{(z_{s3}-z_r)}{\sqrt{(\vec{s}_3-\vec{r})^2}} & 1 \\ -\frac{(x_{s4}-x_r)}{\sqrt{(\vec{s}_4-\vec{r})^2}} & -\frac{(y_{s4}-y_r)}{\sqrt{(\vec{s}_4-\vec{r})^2}} & -\frac{(z_{s4}-z_r)}{\sqrt{(\vec{s}_4-\vec{r})^2}} & 1 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ c\Delta b \end{bmatrix} + \begin{bmatrix} \Delta\varepsilon_1 \\ \Delta\varepsilon_2 \\ \Delta\varepsilon_3 \\ \Delta\varepsilon_4 \end{bmatrix}$$

The fractions in the first matrix are the cosine angles of the unit line of sight from the receiver to the jth satellite. The $\Delta\varepsilon_j$ are the residual errors left after the known offsets have been removed in each pseudorange. If there are additional satellites in view, additional rows are inserted in the above matrix equation for an overdetermined system of equations. If the residual errors are negligible compared to the matrix multiplication and considered zero, then the matrix solution becomes:

$\Delta\mathbf{R} = \mathbf{G}\,\Delta\mathbf{X}$ or $\Delta\mathbf{X} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\Delta\mathbf{R}$ in least-squares solutions, or,

$\Delta\mathbf{X} = (\mathbf{G}^T\mathbf{W}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{W}\Delta\mathbf{R}$ for weighting individual observations with $\mathbf{W}$, and

$\mathbf{X} = \mathbf{X}_o + \Delta\mathbf{X}$ is the navigation solution using this standard method.

The typical approach is to assume at initialization that the nominal position is at the center of the Earth (0, 0, 0), and the navigation solution will converge since the GPS satellites are overhead. Thereafter, the solved position from this last process is used as the next nominal position for $\mathbf{X}_o$. The GPS literature has not had problems with this technique yet, so the opinion has been that it will always work. However, problems with the standard method have not yet been well documented in the literature, specifically that this standard method can converge prematurely, especially when the receiver may be located near a satellite. The problem may occur more often in the real world when more

pseudolites, which are GPS ground transmitters, are installed at airports for additional coverage. Whenever a plane starts up at the airport with a nearby pseudolite, the onboard GPS receiver may converge prematurely to an incorrect navigation solution. This is discussed in further detail in the research portion of this thesis.

5.2 Newton's Method for Systems of Equations

For a system of nonlinear equations **f(x)**, apply Newton's method by taking only the first order partial derivatives about a set of nominal values placed in vector form as **X**.

$$y_1 = f_1(a_1,...a_n) + \frac{\partial f_1}{\partial x_1}(x_1 - a_1) + ... + \frac{\partial f_1}{\partial x_n}(x_n - a_n)$$

$$\vdots$$

$$y_n = f_n(a_1,...a_n) + \frac{\partial f_n}{\partial x_1}(x_1 - a_1) + ... + \frac{\partial f_n}{\partial x_n}(x_n - a_n)$$

This is symbolized as an equivalent one matrix-vector equation.

$$\mathbf{y} = \mathbf{f(a)} + \mathbf{J(a)(x - a)}$$

In Newton's method for a system of equations, the nominal values are initially symbolized as $\mathbf{x_1}$. Replace each component of **f** by its tangent plane approximation at **x** = $\mathbf{x_1}$.

$$\mathbf{y} = \mathbf{f}(\mathbf{x_1}) + \mathbf{J}(\mathbf{x_1})(\mathbf{x} - \mathbf{x_1})$$

The next approximation $\mathbf{x_2}$ to the system is the root at the point where each tangent vanishes or equivalently where **y** is zero. Replacing **y** by zero gives an equation for $\mathbf{x_2}$.

$$\mathbf{0} = \mathbf{f}(\mathbf{x_1}) + \mathbf{J}(\mathbf{x_1})(\mathbf{x_2} - \mathbf{x_1}) \text{ or}$$

$$\mathbf{x_2} = \mathbf{x_1} - \mathbf{J}(\mathbf{x_1})^{-1}\mathbf{f}(\mathbf{x_1})$$

Now, linearize about $\mathbf{x_2}$ to get $\mathbf{x_3}$. Continue this until the desired convergence is achieved.

$\mathbf{x_{k+1}} = \mathbf{x_k} - \mathbf{J}(\mathbf{x_k})^{-1}\mathbf{f}(\mathbf{x_k})$ where the Jacobian is the n x n matrix of

$$\mathbf{J(x)} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_n}{\partial x_n}(\mathbf{x}) \end{bmatrix}$$

Nothing in Newton's method precludes additional convergences from additional solutions whenever the inverse Jacobian is nearly zero from numeric combinations of the partial differentials. Such occurrences are rare as found in this research, but these exceptions do satisfy the usual convergence criteria to stop iterating after n arbitrary attempts even though the desired solution has not been reached. The premature stalling is documented with two regions although other such regions exist in the example given in Chapter 8, which also contains the new exact derivation of the navigation solution.

5.3 Least Squares for an Overdetermined Linear System

Consider a general overdetermined linear system $\mathbf{Ax} = \mathbf{b}$ of m equations with n unknowns as $m \geq n$. The residual is $\mathbf{r} = \mathbf{b} - \mathbf{Ax}$ and the Euclidean norm of the residual is the square root of $\mathbf{r}^T\mathbf{r}$. The least squares solution to $\mathbf{Ax} = \mathbf{b}$ minimizes

$$\mathbf{r}^T\mathbf{r} = (\mathbf{b} - \mathbf{Ax})^T(\mathbf{b} - \mathbf{Ax})$$

The minimum exists for the above matrix equation where the partial derivatives with respect to each of the variables $x^1, x^2, \ldots x^n$ is then set to zero. Differentiate with respect to each variable and set the derivative to zero. Observe that

$$\mathbf{r}^T\mathbf{r} = r_1^2 + r_2^2 + \cdots + r_m^2$$

where r_i is the *i*th component of $\mathbf{r}$. Differentiating the above equation with respect to x_1,

$$\frac{\partial}{\partial x_1}\mathbf{r}^T\mathbf{r} = 2r_1\frac{\partial}{\partial x_1}r_1 + 2r_2\frac{\partial}{\partial x_1}r_2 + \cdots + 2r_m\frac{\partial}{\partial x_1}r_m$$

The *i*th component of $\mathbf{r} = \mathbf{b} - \mathbf{Ax}$ is equal to b_i minus the *i*th row of $\mathbf{A}$ times $\mathbf{x}$:

$$r_i = b_i - a_{i1}x_1 - a_{i2}x_2 - \cdots - a_{in}x_n$$

As the partial derivative of r_i with respect to x_1 is $-a_{i1}$, then the partial equation would be:

$$\frac{\partial}{\partial x_1}\mathbf{r}^T\mathbf{r} = -2r_1a_{11} - 2r_2a_{21} - \cdots - 2r_ma_{m1}$$

In general, the partial derivative of $\mathbf{r}^T\mathbf{r}$ with respect to x_j is given by

$$\frac{\partial}{\partial x_j}\mathbf{r}^T\mathbf{r} = -2r_1a_{1j} - 2r_2a_{2j} - \cdots - 2r_ma_{mj}$$

The right side of the above partial equation with respect to x_j is –2 times the product between the *j*th column of $\mathbf{A}$ and $\mathbf{r}$. The *j*th column of $\mathbf{A}$ is the *j*th row of $\mathbf{A}^T$. Since the *j*th component of $\mathbf{A}^T\mathbf{r}$ is equal to the *j*th column of $\mathbf{A}$ times $\mathbf{r}$, the partial derivative of $\mathbf{r}^T\mathbf{r}$

with respect to x_j is the *j*th component of the vector $-2\mathbf{A}^T\mathbf{r}$. The Euclidean norm of the residual is minimized at that point **x** where all the partial derivatives vanish:

$$\frac{\partial}{\partial x_1}\mathbf{r}^T\mathbf{r} = \frac{\partial}{\partial x_2}\mathbf{r}^T\mathbf{r} = \cdots = \frac{\partial}{\partial x_n}\mathbf{r}^T\mathbf{r} = 0$$

Since each partial derivative is –2 times the corresponding the $\mathbf{A}^T\mathbf{r}$ component, then $\mathbf{A}^T\mathbf{r} = 0$. Replace **r** by **b** – **Ax** and obtain $\mathbf{A}^T(\mathbf{b} - \mathbf{Ax}) = \mathbf{0}$, or in the normal equation form, $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$. Any **x** that minimizes the Euclidean norm of the residual **r** = **b** – **Ax** is a solution to the normal equation. Also, any solution to the normal equation is a least squares solution to the overdetermiend linear system **Ax** = **b**, which makes the least-squares a minimal solution.

Another way of analyzing the least squares is to use vectors. The set of vectors **Ax** corresponding to various choices of **x** is the range space of **A**. The collection of residuals **b** – **Ax** is the set of vectors, which point from the range space to **b**. The shortest residual is perpendicular to the range space. This means the dot product between each vector in the range space and the shortest residual is zero. Since each column of **A** is contained in its range space, the dot product between each column of **A** and the least squares residual must be zero. The dot product between each column of **A** and the residual **b** – **Ax** is given by the matrix-vector product $\mathbf{A}^T(\mathbf{b} - \mathbf{Ax})$. Therefore, the least squares solution (Hager, 1988) to **Ax** = **b** satisfies the matrix equation $\mathbf{A}^T(\mathbf{b} - \mathbf{Ax}) = \mathbf{0}$ as well as the normal matrix equation $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$.

5.4 Doppler

The numerically controlled oscillator (NCO), which regulates the carrier tracking loop, provides an estimate of the observed frequency shift of the received signal. As the user and satellite have motion, the observed frequency differs from the L1 or L2 frequencies. The Doppler shift, D, between the satellite and receiver is the projection of the relative velocity onto the line of sight $\vec{L}$, multiplied by the transmitted frequency and divided by the speed of light.

$$D_j = -f_{L1}\left(\frac{\vec{v}_s - \vec{v}_r}{c}\right) \cdot \vec{L}$$

An estimate of the Doppler shift will keep the tracking loop on the apparent frequency while undergoing motion.

5.5 Accumulated Delta Range

The signal processor produces the accumulated delta range (ADR) by following the commanded values to the NCO required to maintain lock on the signal. The signal processor keeps track of changes in the observed range to the satellite. A common use of ADR values is to smooth the noisy pseudorange measurements. The main difference between ADR and pseudorange measurements is that an ADR has no set starting value. ADR measurements only keep track of the total change in range, but no known method exists to determine instantly the whole number of carrier cycles between the satellite and the antenna with a single receiver. In Differential GPS (DGPS), it is possible to determine the differences in the cycles between two nearby receivers. For precision DGPS surveying, one receiver is the designated base station located at a previously surveyed benchmark and the remote GPS receiver obtains GPS range and cycle data. In postprocessing, the remote receiver's data are limited to the common view satellites and differences against the base station data. The single difference phase observable is constructed by subtracting the phase measurements simultaneously observed with the same satellite, which cancels virtually all the errors associated with that satellite (mostly the satellite clock difference from GPS absolute time). Subtracting the single-difference phase observable data at the same time, which now eliminates the two receiver clock errors in the equations, generates a double-difference phase observable. If the double-difference data are not noisy, then the total number of cycles can be projected onto the line of sight. Integer possibilities are chosen until one stable solution is left while other possibilities drift away from the stationary receiver. This solution now obtains a very precise distance in cm between the two GPS antenna. All of this involves intensive postprocessing and careful editing of anomalous cycle slips in the data to get the desired result (Leick, 1995).

ADR data are commonly used in single GPS receivers to smooth the noisy pseudorange measurements. One technique uses a weighted average of the code and carrier measurements (Hatch, 1982). Another technique integrates the ADR for the

reference trajectory and then the Kalman filter updates the reference trajectory using the pseudrange data at a slower rate (Hwang and Brown, 1989).

This thesis proposes a novel way to apply the knowledge of the pseudorange accuracy and the dual frequency phase measurements to get a revised range measurement within centimeter precision. This technique does not require any accumulated delta range in the real-time GPS measurements and can be accomplished with a single GPS receiver.

5.6 Satellite Navigation Input

The navigation algorithm requires the satellite position as a minimum to solve for the receiver's position and enough information to correct for the offsets in the pseudorange data. The satellite message broadcast in the transmitted GPS signals contains the correction parameters for the satellite clock offsets and other time related effects. The GPS relativity compensation can be computed from the eccentric anomaly term from the satellite ephemeris or from numerically differentiating the satellite ephemeris to get both satellite position and velocity. The atmospheric delays due to the ionosphere and troposphere are discussed in a later section of this thesis. These corrections are applied before computing the GPS point solution.

5.7 Geometric Dilution of Precision

The Dilution of Precision (DOP) is an approximate "rule of thumb" and is often applied to GPS position accuracy estimates. It is a byproduct of the navigation algorithm, so it is often obtained to indicate the accuracy of the navigation solution. Its main purpose was to choose the satellites that would give the best geometry and provide the smallest error in the navigation solution, especially when the early GPS receivers had far fewer tracking channels and could only monitor 4 GPS orbiting satellites from all GPS satellites in view.

The derivation of the DOP equation rests upon the Gauss-Markov Theorem, which proves that the least squares method has the least variance of the unbiased estimators. That means least squares is the best estimation rule provided these assumptions are met:

(1) The statistical model is linear, where $Y_i = BX + e_i$ for i = 1,…,n.

(2) The noise e_i has a zero expectation. (i.e., $E[e_i] = 0$)

(3) All error sources have the same constant variance. (i.e., $var(e_i) = \sigma^2$)

(4) The covariance between individual noise parameters is zero as $cov(e_i, e_j) = 0$. In general, the errors do not have to be normally distributed. Also, the errors are not necessarily independent, but the errors must be uncorrelated, which is a weaker condition. The linear unbiased estimator of $Y = c_1X_1 + c_2X_2 + \ldots + c_nX_n$, does not require that the graph of Y be a straight line (i.e., the variables X_i may have internal powers > 1 or be periodic, etc.). The requirement is that the best estimator Y be linear relative to some set of X variables so that a matrix representation of $\mathbf{Y} = \mathbf{CX}$ holds with $\mathbf{X}$ as a column matrix that satisfies a linear regression. The reality is that none of these four requirements are met in the actual GPS environment.

First, the range equations contain square roots and quadratics, and these equations are linearized by ignoring the higher order terms of the Taylor series. This means the iteration depends upon a good estimate of the initial position and time bias to get a solution. In fact, this thesis documents that the standard GPS iteration converges prematurely in many regions in a simulated satellite constellation. Some exceptional regions occur near a simulated satellite location, which implies this situation may well happen when pseudolites are placed near airports. This has hardly been discussed at all in the GPS literature.

Second, the noise in the pseudorange is not zero mean and independent of time. In actual stationary GPS surveys, the position moves chronologically around the benchmark in curved patterns, usually with the benchmark at the center after many hours of GPS data are recorded. Using the best models and measurements, the individual errors are not random about some mean over a short time interval. GPS positions also exhibit step jumps in meters when a satellite is dropped or added to the GPS least squares solution.

Third, pseudoranges do not have the same variance at any instant and even the pseudoranges from any particular GPS satellite change over time. For example, the GPS broadcast includes the value of the User Range Accuracy (URA). This is a statistical indicator of the ranging accuracies obtainable with a specific satellite. The URA is a one-sigma estimate of the user range errors in the navigation data for the transmitting satellite.

It includes all errors for which the Space and Control Segments are responsible, but it does not include any errors introduced by the receiver or the atmosphere. While the URA may vary over the given fitted interval, the URA that is reported is the maximum value anticipated over the fitted interval (GPS JPO, 2000). The URA is only one component of the contributing variances, which include ionosphere, troposphere, and satellite position. The total variance of any GPS satellite's ranging data, which statistically is the sum of the individual source variances, is never the same constant or the same over time. Moreover, the satellite position error is an elongated ellipsoid, where typically the average satellite position uncertainty is about 4.5 meters along the orbital path, about 3.2 meters cross track or perpendicular to the orbital plane, and 1.2 meters radial between the satellite and Earth's center (Zumberge and Bertiger, 1996). The real satellite variance is not a three-dimensional sphere, but it has directional probabilities with different values in each axis.

Fourth, all GPS satellites exhibit some form of cross correlation. The Control Segment has its Extended Kalman Filter showing the cross correlation terms are not statistically insignificant, although the cross covariances are usually smaller than the variance values along the diagonal of the covariance matrix. Even then, some systematic unknown effects still exist in GPS. For example, the unknown systematic error called Y-axis bias affects all GPS satellites so that each appears to lag 50 meters behind its predicted orbit after a week, despite using the most complete equations of motion and modeling maintained by the Control Segment.

Any of these facts would invalidate the rigorous proof of the following derivation for the Dilution of Precision (DOP). It is noted that many in the GPS community still consider the DOP equation as truth. At best, the DOP factor is an estimator in the uncertainty of the position in the solution. However, it is not always reliable, since the DOP factor can have a reasonably small value while the iterative algorithm converges prematurely to the wrong location. Another example is that with enough visible satellites scattered above the receiver, the position DOP factor can be less than one, meaning the GPS solved position is supposedly more accurate than the inherent uncertainty in the satellites' own positions. This is never possible in surveying. The whole science of adjusted measurements used in surveying is based on rigorous mathematics, and any site

survey obtained by triangulation from a grid of previously surveyed benchmarks with given position uncertainties has no smaller uncertainty than the grid itself. Even with infinite precision, the new surveyed point is tied to the grid, which has some unknown, but definite, displacement from truth, and that would ensure the new point has at least the same minimum error in its determined location. In any case, it will be shown in the analysis of the errors inherent in the iterative GPS algorithm that the DOP factors are inconsistent or inaccurate in determining the actual error in the solved navigation solution.

5.7.1 Derivation of the Dilution of Precision Equation

For the sake of argument, assume the relationship between the pseudoranges and the receiver's position is linear. The form will be

$\Delta\mathbf{R} = \mathbf{G}\Delta\mathbf{X}$ or $\Delta\mathbf{X} = (\mathbf{G}^{\mathrm{T}}\mathbf{W}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{W}\Delta\mathbf{R}$ by applying least squares where the weight matrix W is

$$W = \mathrm{Diag}[\sigma_1^2, \sigma_2^2, ..., \sigma_n^2] \text{ with}$$

$$\sigma_i^2 = \sigma_{\mathrm{Sat\,Pos}}^2 + \sigma_{\mathrm{iono}}^2 + \sigma_{\mathrm{tropo}}^2 + \sigma_{\mathrm{URA}}^2 + \sigma_{\mathrm{low\,elev}}^2 \text{ for } 1 \le i \le n.$$

For brevity, let $\mathbf{K} \equiv (\mathbf{G}^{\mathrm{T}}\mathbf{W}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{W}$. Estimate how close the solution is to truth by getting

$$\mathrm{Cov}(\Delta\mathbf{X}) = \mathrm{E}\left[(\Delta\mathbf{X})(\Delta\mathbf{X})^{\mathrm{T}}\right] = \mathrm{E}\left[(\mathbf{K}\Delta\mathbf{R})(\mathbf{K}\Delta\mathbf{R})^{\mathrm{T}}\right]$$

With $\mathrm{E}\left[(\Delta\mathbf{R})(\Delta\mathbf{R})^{\mathrm{T}}\right] = \mathrm{Cov}(\Delta\mathbf{R})$ = the covariance matrix of the measurement vector $\Delta\mathbf{R}$,

$$\mathrm{Cov}(\Delta\mathbf{X}) = \mathbf{K}\mathrm{Cov}(\Delta\mathbf{R})\mathbf{K}^{\mathrm{T}}$$

The receiver only has no knowledge of the cross correlation that may exist between pseudoranges from different GPS satellites, so the assumption that the measurements are uncorrelated must be employed here. Furthermore, the receiver only has the transmitted information of the URA and can approximate the other variances to estimate the total variance for each ranging measurement. These variances can be set up in the matrix form as **W**, which is an estimate of the Cov($\Delta\mathbf{R}$). This formulation now obtains

$$\mathrm{Cov}(\Delta\mathbf{X}) = \mathbf{K}\mathbf{W}\mathbf{K}^{\mathrm{T}}$$

The derivation so far has only assumed the measurements are unbiased linear estimators of the position and that the errors in the measurements are zero mean and uncorrelated. Intrinsically, the satellite position errors are taken as spherical so that there is invariance due to any rotation inherent in the inversion of the basis where the measurement space is taken and transformed into the 3 dimensional position space. At this point, the covariance is a general measure of the position error in the solution, since the pseudoranges have separate values. However, if $\mathbf{W} = \sigma^2\mathbf{I}$, which implies all measurement weights are equal by being a constant times the identity matrix, then the covariance matrix reduces easily to $(\mathbf{G}^T\mathbf{G})^{-1}$. Note a rotation matrix is a square unitary matrix where $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. So, the covariance basis can be rotated from the ECEF XYZ frame to the local level coordinates of east, north and up by multiplying $\mathbf{G}$ with $\mathbf{U}$.

$$\mathrm{Cov}(\Delta\mathbf{X}) = \sigma^2\left(\mathbf{G}^T\mathbf{G}\right)^{-1} = \sigma^2\left(\mathbf{G}^T\mathbf{U}^T\mathbf{U}\mathbf{G}\right)^{-1} = \sigma^2\left(\mathbf{L}^T\mathbf{L}\right)^{-1} = \sigma^2\begin{bmatrix} \sigma^2_{east} & \sigma_{en} & \sigma_{eu} & \sigma_{et} \\ \sigma_{en} & \sigma^2_{north} & \sigma_{nu} & \sigma_{nt} \\ \sigma_{eu} & \sigma_{nu} & \sigma^2_{up} & \sigma_{ut} \\ \sigma_{et} & \sigma_{nt} & \sigma_{ut} & \sigma^2_{time} \end{bmatrix}$$

The Dilution of Precision (DOP) is based on the above equation. The geometric DOP is GDOP, the position DOP is PDOP, the horizontal DOP is HDOP, the vertical DOP is VDOP, and time DOP is TDOP. The various DOP factors are defined in east, north, up and time components as follows:

$$\mathrm{GDOP} = \sqrt{\sigma_e^2 + \sigma_n^2 + \sigma_u^2 + \sigma_t^2}$$

$$\mathrm{PDOP} = \sqrt{\sigma_e^2 + \sigma_n^2 + \sigma_u^2}$$

$$\mathrm{HDOP} = \sqrt{\sigma_e^2 + \sigma_n^2}$$

$$\mathrm{VDOP} = \sigma_u$$

$$\mathrm{TDOP} = \sigma_t = c\sigma_{time}$$

If σ^2 represents each satellite's total variance, the DOP equation predicts the solution error from an estimated pseudorange error of σ will be:

$\delta = \sigma \times \mathrm{GDOP}$ where GDOP is a geometric function of the satellite constellation. As emphasized, this is a crude approximation of the actual navigation errors. For example, any errors that are numerically identical in all pseudoranges will become

mapped into the clock error, because any common range errors look like clock errors as $\delta R = c\,\delta t$. When that happens, the PDOP is small while TDOP is artificially higher than actual clock performance errors. The GDOP equation oversimplifies the navigation errors and the real effects.

The GDOP of an overdetermined satellite geometry is always less than GDOP for any subset of 4 satellites. This is because with each additional satellite, the volume of the polyhedron defined by vertices located at the satellite positions and the receiver position is always larger. GDOP is inversely proportional to this volume (Spilker, 1996b). For example, under the assumption of equal measurement variances, the optimal 4 satellite geometry (3 satellites equally spaced on the local horizon, one satellite at the zenith) has a GDOP of $\sqrt{3}$. The optimal 5 satellite geometry (4 satellites equally spaced on the horizon and 1 satellite at the zenith) has a GDOP factor of $\sqrt{2}$.

5.7.2 Applying Dilution Equation in Local Coordinates

The GDOP is the square root of the trace of the covariance matrix $\mathbf{\Delta X}$ without the scalar σ^2, which is an invariant quantity independent of the coordinate frame. To use the other quantities, such as HDOP in the local horizontal plane, the unitary matrix $\mathbf{U}$ must be obtained to transform the covariances into the local east, north and up components computed from the GPS XYZ frame. To set up the rotation in local east, local north and local up, the X axis would be rotated -93.62223004 degrees about the north pole (Z axis) to be in the local meridian, which is the longitude through Ames. But to get the axis parallel to the local east, the X axis would be rotated +90 degrees further, and the net rotation is -3.62223004° for ψ. The next rotation θ is about the newly placed X axis to rotate the Z axis parallel with the local up vector at Ames, which is 90 degrees – geodetic latitude of Ames or 47.977775975°. The resulting rotation matrix, which is unitary, to transform distance measurements into local coordinates is:

$$\mathbf{U} = \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\cos\theta\sin\psi & \cos\theta\cos\psi & \sin\theta \\ \sin\theta\sin\psi & -\sin\theta\cos\psi & \cos\theta \end{bmatrix} \quad \text{and} \quad \left[(\mathbf{GU})^{\mathrm{T}}\mathbf{GU}\right]^{-1} = \mathbf{\Delta X}$$

Compute the various DOP values according to the formulas listed in the previous section. The results are summarized in the following table.

Local Dilution of Precision Values	
GDOP (enut)	3.22
PDOP (enu)	2.74
HDOP (en)	2.39
NDOP (n)	2.27
EDOP (e)	0.76
VDOP (u)	1.35
TDOP (t)	1.70

Table 1: Dilution of Precision in East, North, Up Components

Even though the coverage is limited to the northern half of the local hemisphere, the DOP values are reasonable, which indicate that navigation solutions should be stable using the standard computational method. Ideally, the GDOP should be as close as possible to 1, and this could only be achieved with additional satellites with some additional satellites in the southern portion of the local hemisphere.

5.8 Earth Rotation Compensation to Pseudoranges

A race's completion time is altered when the finish line is moved relative to the runners during the race, and, similarly, the pseudorange must be compensated for a ground-based receiver due to the rotating Earth. The longest transmission interval exists with the GPS satellite at the local horizon, which is about 25,782,600 meters in distance or 86 ms in duration. The ground-based GPS receiver has the largest rotational velocity of 465 m/s at the equator, and the Earth's rotation displaces that receiver by a distance of 40.0 m in that pseudorange with the satellite due east or west of the receiver. The time of transmission is just the pseudorange, ρ, divided by the speed of light, c. The receiver's rotational velocity depends on its latitude on the spinning Earth, or equivalently, $\boldsymbol{\omega} \times \mathbf{R}$ with $\boldsymbol{\omega}$ being the angular rotation rate in radians per second and $\mathbf{R}$ being the receiver's XYZ location. The velocity multiplied by the transmission interval will obtain the correction to the GPS receiver's pseudorange. This is in vector notation for either form as:

$$(\boldsymbol{\omega} \times \mathbf{R}) \cdot \boldsymbol{\rho} / c^2 = dt \quad \text{or} \quad (\boldsymbol{\omega} \times \mathbf{R}) \cdot \boldsymbol{\rho} / c = c\ dt = d\rho$$

5.9 Relativity

Relativity affects the GPS satellite clock, both in the velocity contribution from special relativity and in the gravitation difference from general relativity in space versus the Earth's surface. Without getting into the detailed relativistic derivations, the collective effects to first order have been derived (Ashby, 1987) to find that:

$$\delta t_{rel} = \frac{f_0^{'} - f_0}{f_0} = \frac{1}{2}\left(\frac{v}{c}\right)^2 + \frac{\mu}{c^2}\left[\frac{1}{R_E + h} - \frac{1}{R_E}\right] = -4.465 \times 10^{-10}$$

To receive this nominal frequency f_0 at the Earth's surface, the GPS satellite clock is offset from the ideal frequency of 10.23 MHz to 10.22999999543 MHz, so that the GPS receiver does not need to make any compensation for these effects. However, the GPS satellite traverses in an elliptical orbit that varies the gravitational potential and the velocity. The GPS receiver compensates for these effects by the following formula (GPS JPO, 2000) (Hoffman et al, 1994):

$$\Delta t_{rel} = -\frac{2}{c}\sqrt{\mu a}\, e \sin E = -\frac{2\vec{P}\cdot\vec{V}}{c^2},$$

where $\vec{P}$ and $\vec{V}$ are the satellite's position and velocity.

Chapter 6 Atmospheric Effects on GPS Pseudoranges

GPS still exhibits significant errors. The largest natural errors come from the atmosphere, and this chapter discusses the largest effects that alter the accuracy of the pseudoranges. Although the tropospheric effects at very low elevations are sometimes comparable with the ionosphere effects, the ionosphere has far greater variability.

6.1 Ionospheric Effects on GPS

The ionosphere is formed basically by ionization from the Sun's radiation. The ionosphere is a weakly ionized plasma gas, which has different chemical properties at different altitudes. The density of the ions is less than 1% of the neutral gases. The various regions of the ionosphere have different chemical properties with the heavier molecules and ions at lower heights and the lighter atomic ions at greater altitudes. Electron diffusion becomes important above 180 km above the surface, and the free electrons that are generated by solar ultraviolet emissions follow the Earth's magnetic lines of force. Additional changes in electron density above 180 km are produced by electric fields that cause electrons to move in a direction perpendicular to the magnetic lines of force. The scale height of each ionic molecule is $H = kT/mg$ where k is Boltzmann's constant, T is temperature in Kelvin, m is the mass of the molecule and g is the gravitational acceleration. The scale height is inversely proportional to each ion's atomic weight, but free electrons do not fall off very fast with increasing height.

The ionosphere produces most of the effects on GPS signals by the total number of free electrons in the ionosphere. This integrated number of electrons is called the total electron content (TEC), and it is expressed as the number of electrons in a vertical column having 1 square meter cross section between the satellite and the receiver.

Fortunately, the ionosphere is dispersive where the refractive index is a function of the operating frequency. The ionosphere range error can vary from only a few meters to many tens of meters during the course of each day. The two-frequency GPS receivers take advantage of this property and correct for the first-order ionospheric range and range-rate effects directly.

6.1.1 Refractive Index of the Ionosphere

To categorize the propagation effects of radio waves through the ionosphere, its refractive index must be specified. The refractive index of the ionosphere, n, has been derived by Appleton and Hartree (Davies, 1989) and is shown

$$n^2 = 1 - \frac{X}{1 - iZ - \frac{Y_T^2}{2(1 - X - iZ)} \pm \left[\frac{Y_T^4}{4(1 - X - iZ)^2} + Y_L^2 \right]^{1/2}}$$

where $X = Ne^2/\varepsilon_o m\omega^2 = (f_n/f)^2$, $Y_L = eB_L/m\omega = (f_H \cos\theta)/f$, $Y_T = eB_T/m\omega = (f_H \sin\theta)/f$, $Z=\nu/\omega$, $\omega = 2\pi f$, where f is the system operating frequency in Hz, and

e = electron charge, -1.602×10^{-19} coulomb

ε = permittivity of free space = 8.854×10^{-12} farad/m

m = rest mass of an electron = 9.107×10^{-31} kg

θ = the angle of the ray relative to the Earth's magnetic field

f_ν = the electron-neutral collision frequency $\approx 10^4$ Hz.

f_H = the electron gyro frequency $\approx$ 1.5 MHz

f_n = the plasma frequency is $\leq$ 20 MHz in virtually all cases.

Using the newly defined terms for X, Z and $Y = \sqrt{Y_L^2 + Y_T^2}$, the refractive index of the ionosphere at GPS frequencies with all coefficients greater than 10^{-9} can be derived (Brunner and Hajj, 1991) as:

$$n = 1 - (X/2) \pm (XY/2)\cos\theta - (X^2/8)$$

The last two terms are far less than the first order term, X/2. For a more typical ionospheric maximum plasma frequency of 12 MHz, the higher order terms can be ignored to better than 0.1% accuracy. Limiting the angle of incidence above 70 degrees (except near the magnetic pole regions), the error in the dual frequency ionosphere compensation is now less than 0.0001 of the ratio between the first and second order terms. Thus, a correction of the first order term of, say, 100 meters will result in a compensation error less than 1 centimeter. The Appleton-Hartree equation reduces to the simple dual frequency ionosphere correction for the refractive index as:

$$n = 1 - (X/2) = 1 - \frac{40.3}{f^2}\int N\, dl = 1 - \frac{40.3}{f^2}\,TEC\,.$$

The quantity $\int N\,dl$ is the total electron content (TEC), which is in units of el/m^2 of free electrons in the volume of the atmosphere integrated along the propagation path between the GPS satellite and receiver. One TEC unit = 1 x 10^{16} el/m^2, which causes 0.853 cycles of phase in L1 or 0.163 m range error in L1. With the current GPS measurement precision, the dual frequency ionosphere compensation is quite sufficient for all practical GPS applications to less than 1 centimeter.

6.1.2 Ionospheric Group Delay

The delay of the ionosphere produces range errors, which can be expressed as time delay or, in distance, by multiplying the time delay by the speed of light. This delay can be determined by $\Delta r = \int (1 - n)\,dl$,. At L band, the first-order refractive index is $n = 1 - X/2$ where $X = 40.3(\int N\,dl)/f^2$ and the ionospheric delay is $\Delta t = \frac{40.3}{cf^2}\int N\,dl$.

Let I_i be the ionosphere effect for the ith frequency, either f_{L1} or f_{L2}, D be the corrected propagation distance between the satellite and receiver, and PR_i be a measured pseudorange of the ith link. Since the geometric range is the same from any satellite to the receiver antenna, the difference of the simultaneous measurements of the total range will give the effective value of the TEC. Pseudoranges have the same time offsets due to the receiver's internal clock and due to other offsets common to both frequencies, such as the troposphere. Thus, the ionosphere's time delay difference is observed by the difference between the measured pseudoranges.

$$D = PR_1 - I_1 = PR_2 - I_2 \text{ where } I_1 = \frac{40.3}{f_1^2}\text{TEC.}$$

$$I_2 - I_1 = PR_2 - PR_1 = c\,\delta t = 40.3\,\text{TEC}\left(\frac{1}{f_2^2} - \frac{1}{f_1^2}\right) \text{ where by design, } f_1 = \frac{77}{60} f_2.$$

$$\delta t = (40.3/c) \times \text{TEC} \times \left[\frac{1}{f_2^2} - \frac{1}{f_1^2}\right] = \Delta t_1 \left[\frac{f_1^2 - f_2^2}{f_2^2}\right] \text{ as } \Delta t_1 = I_1/c \text{ or}$$

$$\Delta t_1 = \left[\frac{f_2^2}{f_1^2 - f_2^2}\right] \times \delta t = 1.546 \times \delta t = 1.546 \times \frac{(PR_1 - PR_2)}{c}$$

Here, L1 frequency is f_1, L2 frequency is f_2, and the measured time delay difference is δt. To put this in terms of corrected PR = $PR_1 + c\,\delta t$, which by substitution is:

$$D = \text{Corrected PR} = PR_1 + \left[\frac{f_2^2}{f_1^2 - f_2^2}\right] \times (PR_1 - PR_2) = \frac{PR_2 - \frac{77^2}{60^2} PR_1}{1 - \frac{77^2}{60^2}}$$

$D = PR_2 - I_2$ also checks with the above equation.

6.1.3 Ionospheric Carrier Phase Shift

The carrier tracking loop inside the GPS receiver counts how many cycles have been generated per unit time. Since the GPS signal is delayed due to the ionosphere, both the code and carrier are delayed, but the carrier tracking loop is still generating cycles to compare to the input signal. The code chip boundary is set by 180 degree changes in the carrier phase when transitioning between values of 1 to -1 or -1 to 1. The carrier phases at code boundaries appear as normal phases when the code chip value remains the same (-1 to -1, or 1 to 1). When the delayed signal does arrive, the carrier tracking loop has precisely more cycles when it compares the internally generated carrier to the input carrier, and the result is an apparent advance in the integrated carrier phase measurement. This carrier phase advance due to the ionosphere can be expressed as:

$$\Delta\phi = \frac{1}{\lambda}\int (1+n)dl \text{ in cycles or wavelengths to be subtracted later, or}$$

$$\Delta\phi = -f\,\Delta t = -f \times \left[\frac{f_2^2}{f_1^2 - f_2^2}\right] \times \frac{(PR_1 - PR_2)}{c}$$

The TEC term is replaced with the value obtained from the measured time delays. This obtains the following correction terms:

$$\text{Corrected } \phi_{L1} = \phi_{L1} + \Delta\phi_{L1} = \phi_{L1} - \frac{f_1 f_2^2}{f_1^2 - f_2^2} \frac{(PR_1 - PR_2)}{c}$$

$$\text{Corrected } \phi_{L2} = \phi_{L2} + \Delta\phi_{L2} = \phi_{L2} - \frac{f_2^3}{f_1^2 - f_2^2} \frac{(PR_1 - PR_2)}{c}$$

These correction equations applied to the pseudorange are only valid if the pseudorange error is insignificant. But, the general pseudorange error is between 20 and 55 centimeters for good signal conditions, and the L1 and L2 scaling factors of 1.546 and

2.546 make multiple cycles of carrier phases. Thus, the carrier phases will not be corrected for this study using the dual frequency ionosphere method, and only pseudoranges will be given the ionosphere correction. As the novel method is illustrated, the ionosphere correction applied to the carrier phases is not needed.

Other ionospheric effects in the range of the GPS signal spectrum are usually insignificant except for the most rigorous applications. Since the frequency is simply the time derivative of phase, the additional Doppler shift due to changing TEC is small and around 0.085 Hz at L1, which corresponds to a 1.5 cm/s range rate error. This effect is insignificant for GPS receivers with a typical tracking loop bandwidth of a few Hz.

The Faraday polarization rotation is overcome in GPS, since the GPS signals are transmitted with right-hand circular polarization. The GPS receiver antenna is designed for right-hand circular polarization to ensure that it matches the characteristics of the received signal. The Faraday polarization rotation affects those transmissions that are elliptically polarized and the antenna is not configured to receive these signals, but the GPS signal is nearly circular with nearly equal power upon incidence with the antenna.

The ionosphere can refract the GPS signal by an angular deflection that is a function of TEC. The findings of low local elevation only resulted in a 4 arcsecond deflection with L1 and 6 arcsecond deflection with L2 at 0 degrees (Millman and Reinsmith, 1974). GPS receivers should not track GPS satellites below 5 degree elevations, since multipath (i.e. reflections) and tropospheric effects are greatly increased at low elevations. So, ionospheric refraction is not significant to GPS.

The ionosphere can produce dispersion of the spread spectrum signals from GPS. The pulse dispersion in differential time delay is proportional to TEC and inversely proportional to the cube of the frequency (Millmann, 1965). The effect across the 20 MHz GPS bandwidth is nearly the usual ionospheric effect /frequency and can be ignored for all practical purposes.

Irregularities in the Earth's ionosphere demonstrate short-term signal fading. These are brief periods, often around the times of maximum 11-year solar cycle period, which strong scintillation effects have been observed. The areas that are hardest hit are near equatorial regions and are generally limited to approximately 1 hour after local

sunset to local midnight with a maximum loss of about 20 dB in signal to noise ratio (Goodman and Aarons, 1990). Severe phase scintillation can cause a large random fluctuation superimposed upon the normal Doppler shift due to the rapid change in TEC. GPS receivers may have phase lock problems, especially in a phase lock loop (PLL) if the ionosphere produces a phase change faster than the receiver's bandwidth can allow. A change of 1 radian in phase at L1 corresponds to 0.19 TEC units, or only 0.2% of the typical ionosphere background, which cause problems to maintain receiver carrier lock loop. To quantify the possible shift, the fluctuation in daytime zenith TEC was recorded at Hamilton, MA, during 1981, a year of high solar activity. The worst deviation of TEC was approximately 25 units over a 1 minute interval. That equates to 21 revolutions in phase over 60 seconds (i.e. phase changes of 125 degrees per second). However, the frequency lock loop (FLL) is able to follow multiple rotations of the phase without loss of lock, so the basic FLL with a bandwidth of 5 to 8 Hz should be able to handle this extreme phase shift. Those regions of Earth with strong phase scintillation are limited normally to near-equatorial latitudes, but during the major magnetic storms of March 1989, these effects occurred over the mid latitudes as well.

Most GPS references claim that the ionosphere retards the code by an amount below the speed of light in a vacuum while the phase is advanced by the same magnitude above the ideal speed of light. For example, Klobuchar (1996) states, "As a radio signal traverses the ionosphere, the phase of the carrier of the radio frequency transmission is advanced from its velocity in free space….The relationship between group delay and carrier phase is simply $\Delta\phi = -f\,\Delta t$, or, for every cycle of carrier phase advance, there are 1/f seconds of time delay." Lieck (1995) points out, "The phase velocity is larger than vacuum speed and the group velocity is smaller than vacuum speed by the same amount Δc." The GPS receiver removes the initial Doppler shift in the transmitted signal and despreads it, there is only one frequency being tracked. It is the phase angle that gets reversed frequently to mark the edges of the code, but the phase velocity remains unchanged with demarking code boundaries by phase reversals.

Assume the GPS transmission is high enough in signal strength so that spreading and dispreading of the signal is unnecessary to recover the standard GPS signal that is

weaker than thermal background noise. The code is physically attached to the carrier and the code boundaries are demarked by transitions of 180° shifts in the phase.

Once the GPS signal is transmitted, there is no way to separate the code from the phase since it is a single frequency. The phase observable is the difference between the generated phase inside the receiver compared to the received phase traversing through the ionosphere. If the transmitted phase and code are equally delayed through the ionosphere, the phase difference is greater since the receiver has generated more phases before the received GPS phase signal arrives. This phase difference is the real advanced measurement in the phase tracking loop and is not the result of the incorrect justification that the phase velocity is larger than the speed of light in a vacuum.

The code is not an envelope of superimposed signals. For example, one code chip is like a step function in baseband. A Fourier summation of frequencies from $-\infty$ to $+\infty$ in time is required to equal a single step function of finite width. This is unrealizable in practice, since frequencies have to be present prior to the existence of the step function. Also, the signal passes through a low pass filter in the receiver, so all high frequencies are effectively cut off. This also means the group envelope is impossible to construct by superpositioning, since all higher frequency components needed to construct the step function mathematically are eliminated.

The phase advance explanation is incorrectly applied to GPS. Undergraduate textbooks on electromagnetic theory document that the phase and group velocities through a medium or even in a hollow waveguide satisfy the equation, $v_g\, v_p = v_c\, v_c$ in a medium such as gas (Lorraine and Corson, 1970). Here, the equation incorporates the index of refraction as slightly larger than 1 as the speed of light through the air is less than in a vacuum. Assume the claim that the code is slower than v_c by δv for the group velocity and the phase is faster than v_c by δv. Directly substitute these expressions into the equation to get a contradiction:

$$v_g v_p = (v_c - \delta v)(v_c + \delta v) = v_c^2 - \delta_c^2 < v_c^2$$

Obviously, the standard explanation is wrong while still getting the right answer from the observed phase measurements. The alternate explanation given here still

obtains the right results, but the explanation offered in this thesis does not contradict electromagnetic theory.

6.2 Tropospheric Effects on GPS

This section deals with the tropospheric effects on the GPS signals and the effects on the GPS measurements. Primarily, the effects discussed include troposphere attenuation, scintillation and delay. It should be noted that about a fourth of these effects are actually attributed to the gases in the tropopause and stratosphere. The troposphere produces attenuation mostly below 0.5 dB and delay effects generally between 2 and 25 meters. These effects vary with elevation, since a longer path length traverses through the troposphere at lower elevation angles.

6.2.1 Troposphere Components Affecting GPS Signals

The troposphere consists of dry gases and water vapor, which affect the propagation delay of the radio frequency signals quite differently. As water vapor is highly variable throughout the atmosphere, the standard method is to model or measure the total amount of water in a vertical column of air with a cross section of one square cm. The total integrated water content can vary widely from the polar region to the equator.

The dry atmosphere is relatively uniform in its content. The main component that attenuates the GPS signals the most is oxygen. The attenuation is about 0.035 dB for a satellite at zenith (Boithias, 1982) (Spilker, 1995). However, the effect can be 10 times larger in dB at very low elevations. Assuming a spherical uniform shell of height H_m above the Earth, the obliquity or projection factor via the troposphere is a function of the elevation angle E for signal strength A in dB (Spilker, 1996a).

$$A(E) \approx \frac{2A_{90\text{deg rees}}(1+a/2)}{\sin E + \sqrt{\sin^2 E + 2a + a^2}} \approx \begin{cases} \dfrac{2A_{90\text{deg rees}}}{\sin E + 0.043}\,\text{dB as } 10° > E > 3° \\ \text{or } \dfrac{A_{90\text{deg rees}}}{\sin E}\,\text{dB for } E > 1 \end{cases}$$

It should be noted that below 3 degrees, the above equation should not be used since the spherical uniform model is not accurate for the troposphere. In practice, the GPS receivers should avoid using satellites below 5 degree elevations.

6.2.2 Tropospheric Delay

The received GPS signal is refracted by the atmosphere, which causes a slightly curved path during transmission. Part of the delay is the actual increase in the path length that the signal travels to reach the antenna. Most of the delay is due to the larger refractive index of the atmospheric gases than for free space. The difference between the actual total path delay $S = c\tau$ and the geometrical shortest path distance S_g is the excess tropospheric delay Δ. The difference between the actual refractive index and unity is $n - 1 \approx 2.7 \times 10^{-4}$ at sea level and varies with altitude, latitude and weather. Both the group velocity and phase velocity are the same, so the troposphere is nondispersive as opposed to the ionosphere.

As the refractive index varies, the actual signal path has a slight curvature with respect to the geometric straight line. The total length of the actual path from P_1 to P_2 is:

$$S = \int_{actual} ds > S_{geometric} \quad \text{and} \quad \tau = \int_{actual} \frac{n(s)\,ds}{c}$$

However, Fermat's principle of least action states that the actual path minimizes the time elapsed (or time delay as $c\tau$) from P_1 to P_2 (Sears, 1949). The actual curved path length S is longer than the straight line geometric path in meters where:

$$S_g = S_{geometric} \approx \int_{geometric} ds \text{ and } c\tau_{geo} = \int_{geometric} n(s)ds$$

The quantity of primary interest is the quantity of excess delay caused by the troposphere as $\Delta = c\tau - S_g$. The error caused by the neglect of path curvature is less than 3 mm for elevations above 20 degrees, 2 cm at 10 degree elevation, but 17 cm at 5 degree elevation. The assumption of concentric spherical shells of thickness δr_i with atmospheric density and with constant refractive index n_i in each shell obtains the spherical form of Snell's law.

$$r_u n_u \sin\psi_u = r_1 n_1 \sin\psi_1 = n(\text{zenith}) R_{radial} \sin\psi(\text{zenith}) = \text{constant}$$

Summarizing, the curved path is physically longer than the straight line, but the curved path has a shorter total time delay than the straight line path. Thus, the troposphere time delay is:

$$\Delta = c\tau - S_{geometric} = \int_{actual} n(s)ds - \int_{geometric} ds$$

$$= \int_{actual} (n-1)ds + \left[\int_{actual} ds - \int_{geometric} ds\right]$$

$$= \int_{actual} (n-1)ds + \Delta_g = (10^{-6})\int_{actual} N ds + \Delta_g$$

where the refractivity $N \approx (n\text{-}1) \times 10^6$ has both dry and wet terms. The excess delay Δ can be expressed as the sum of two contributions from the dry air effects and the wet vapor effects. For practically, troposphere attenuation (typically 0.035 dB at zenith to 0.38 dB at 5 degree local elevations), rainfall attenuation (about 0.081 dB for 20 degree elevation versus 0.259 dB at 5 degree elevation), and troposphere scintillation are insignificant (Spilker, 1996a).

6.2.3 GPS Receiver Troposphere Model

The present troposphere model used in most GPS receivers is the Central Radio Propagation Laboratory (CRPL) Reference Refractivity Atmosphere–1958 (Hotovec, 1980). The US government systems specification for GPS is the document SS-US-200, which requires that the uncorrelated error attributable to troposphere delay shall have a standard deviation error of less than 2 meters. The correlated or constant error of the CRPL 1958 model has been determined as 0.0824 meters from local zenith. The lowest elevation acceptable is E = 5° above the local horizon. The obliquity equation from zenith can be used to determine the worst troposphere error (ET), which then is:

$$E_T = \frac{E_C}{\sin 5° + \dfrac{0.00143}{\tan 5° + 0.0455}} = 0.8414 \text{ meters.}$$

One standard deviation of uncorrelated error (EU) will be EU = 0.8414 – 0.0824 = 0.759 meters. This is the main reason for choosing this model for GPS receivers with no wet air terms, since it easily meets the SS-US-200 requirement

The index of refraction, n, is the ratio of the speed of light in a vacuum, c, to the speed of light in the medium, v. For the atmosphere, $v \approx c$, and the difference shows up

in the third decimal place. In practice, another index is chosen, N, that is independent of frequency to 30 GHz.

$N = [(c/v) - 1] \times 10^6 = [n - 1] \times 10^6$, which is used in the CRPL 1958 model below:

$$R(h) = R_1 + R_2 + R_3 \quad \text{for range corrections in millimeters at height h above sea level}$$

$$R_1(h) = \left[N_s\, h + 0.5\, \Delta N\, h^2 \right]_0^{1\,km} \quad \text{except } R_1 = 0 \text{ when } h_{UE} \geq 1\,km$$

$$R_2(h) = \left[\frac{8\, N_1}{\ln(N_1/105)} e^{[0.125(1-h)\ln(N_1/105)]} \right]_{1\,km}^{9\,km} \quad \text{except } R_2 = 0 \text{ when } h_{UE} \geq 9\,km$$

$$R_3(h) = \left[\frac{105}{0.1424} e^{(1.2816 - 0.1424h)} \right]_{9\,km}^{h_{SV} \approx 20182\,km}$$

where:

h = altitude above sea level of user equipment in km

N_s = refractivity at sea level = 324.8

N_1 = refractivity at one kilometer = $N_s + \Delta N$

$\Delta N = -7.32\, e^{[0.005577\, N_s]}$

This R(h) is the zenith range correction. At other elevations than 90°, use the obliquity equation:

$$\Delta R = \frac{R(h)}{\sin E + \dfrac{0.00143}{\tan E + 0.0455}}$$

for the troposphere correction applied to pseudoranges.

6.2.4 Other Empirical Troposphere Models

Several troposphere models incorporate inputs for changing conditions in the wet and dry components of the atmosphere. About 90% of the tropospheric refractions arise from the dry and about 10% from the wet component (Janes et al, 1989). The following are brief descriptions of various troposphere models in the literature providing slightly

more accurate troposphere models. The reader can refer to the cited articles for details of the models and applications.

The first is the Hopfield Two Quartic model, which has two quartic polynomial fits for the dry and wet components of the troposphere (Hopfield, 1969 and 1970). Using real data covering the whole Earth, Hopfield empirically found a representation of the dry refractivity as a function of the height h above the surface by assuming a single thick polytropic layer:

$$N_d^{Trop}(h) = N_{d,0}^{Tropo}\left[\frac{h_d - h}{h_d}\right]^4.$$

Due to a lack of an appropriate alternative, the Hopfield model assumes the same functional model for the wet components. The input wet values are based on the relative humidity taken by hydrometers at the ground station, usually where the GPS user is nearby. The dry components are modified by the pressure and temperature inputs at the station. Instead of the above empirical function, lengths of the position vectors at sea level, R_E, the user at R and the satellite, R_d, replace the heights h and h_d. This modified Hopfield model is then used to derive a mapping function in the same form as the obliquity equation.

The standard Saastamoinen delay model incorporates the dry pressure using a constant lapse rate for the troposphere and an isothermal model above the tropopause plus a wet refraction factor dependent on the wet partial pressure that decreases more rapidly than the total pressure (Saastamoinen, 1972 and 1973). The Saastamoinen model uses parameters for height corrections in meters and pressure adjustments in millibars for the apparent zenith and station heights above sea level in a large lookup table. An extended Saastamoinen model (Bauersima, 1983) refined this model by adding two correction terms, one being dependent on the observing site's height and the other on the height and zenith angle.

The Black and Eisner model began as a correction model from point P1 to point P2 at various elevation angles. Later, it was modified to ray bending through the troposphere. The dry index of refraction is based on the zenith delay from numerical

integration of the quartic dry formula in the Hopfield model. The wet index is an exponential decay from saturated surface conditions.

Berman developed a simple model (Berman, 1976) for water vapor zenith delay based on day and night measurements at Edwards AFB, CA. By estimating the dry zenith delay from surface pressure measurements and estimating the surface wet and dry refractivity, another estimate of the wet zenith delay is obtained as a ratio of the wet and dry surface refractivity where the scale factor is determined from day, night or mixed periods of time.

Mapping functions can accurately relate actual excess path delay as a function of elevation angle and atmospheric conditions compared to the zenith excess delay. These are similar to the obliquity equations shown above. For angles near 90°, the mapping function can be reduced to 1/sin E. At lower elevations, the finite spherical shell width and nonuniform density for each shell proves to be inadequate. Marini (1972) described a continued fraction with constants for the first few levels of the infinite continued fraction. Chao (1974) had two separate mapping functions for wet and dry components with different constants in place of the constants of the obliquity equation. Davis (1985) had a third level continued fraction for his mapping function that incorporated both dry and wet components after computing the coefficients, which were dependent on the pressures, height, temperatures, and tropospheric temperature lapse rate. Other mapping functions have been derived often based on the modified Hopfield model, such as Yionoulis (1970), Goad and Goodman (1974), Herring (1992), and Black (1978). Many other tropospheric models are similar to those already listed [e.g. Lanyi (1984), Marini and Murry (1973), Elgered (1985), Rahnemoon (1988)], but this is by no means complete.

Many comparisons have been made between these numerous troposphere correction models. One paper (Mayer et al, 2000) examined the differences between 6 models (i.e. Extended Saastamoinen, Standard Saastamoinen, Modified Hopfield, Niell, Herring, and Davis). The conclusion was that the Modified Hopfield, Niell, Herring and Davis models had nearly identical precision, while the Standard Saastamoinen was 1.3 times worse and the Extended Saastamoinen was 1.8 times worse. The relative residuals

after double differencing between two Antarctic locations showed the troposphere models still varied between -4 to +4 centimeters using moving averages, although exceptional trends that were significant still are left unexplained by any troposphere model.

6.2.5 GPS Control Segment Troposphere Model

The GPS Control Segment uses its own model to compute the excess tropospheric delay for corrected pseudoranges. The overall equation for tropospheric delay is:

$$\Delta(E) = \frac{0.02312\, P_s[T - 4.11 + 5(r_m - r_a)/148.98]}{T\sqrt{1 - \left[\frac{r_a \cos E}{r_m + (1-C)(r_d - r_m)}\right]^2}} + \cdots$$

$$\cdots + \frac{0.0746 eh[1 + 5(r_m - r_a)/h]}{T^2\sqrt{1 - \left[\frac{r_a \cos E}{r_m + (1-C)h}\right]^2}} \text{ as } E > 5°$$

The terms are: P_s is the monitor station barometric pressure in kilopascals, T is the monitor station measured temperature converted to Kelvin, r_m is the radial distance from the Earth's center to the meteorological sensors, r_a is the radial distance from the Earth's center to the monitor station antenna, r_d is the tropospheric dry radius, E is the satellite elevation, h is the wet height of the troposphere, e is the estimated partial water vapor pressure, and C = 0.85 as the integration constant. This GPS troposphere delay model is for ground stations in the Control Segment to monitor atmosphere conditions in real time while simultaneously collecting GPS data in normal operations.

Chapter 7 Exact Linear Solution Derived for GPS Navigation

The standard GPS method assumes a minimum of four noncoplanar GPS satellites are all that are needed to get a trilateralized solution from distances only. It actually requires five satellites to get a unique three dimensional position and to get the time bias between the GPS timescale and the receiver's clock. A quick example would prove this assertion. Any three satellites will define a plane. If three ranges are extended from each satellite to a common point representing the receiver's location, then there are two possible pyramids, each sharing a common base in the plane of the three satellites. One pyramid is a mirror image of the other, and the two apexes are the two possible receiver locations. It takes a fourth noncoplanar satellite with a realistic range that terminates at only one of the two possible locations to resolve which position is real and which is extraneous. Since there is an unknown time bias still unresolved, it takes a fifth satellite to solve for the position and the clock difference. The reason the iterative method works with a minimum of four satellites is that the time bias is assumed to be a fixed value (usually zero initially or the last calculated time bias) before the next iteration.

The following derivation requires a minimum of five satellites to solve for the position in XYZ coordinates (ECEF) and the time bias B simultaneously in one operation. Let R_i represent the range between the origin of the ECEF frame and the GPS satellite, and let R_r be the distance between the origin and the receiver. Let D_i represent the geometric distance between the *i*th satellite and the receiver. Let the time bias $B = c(t_{GPS} - t_r)$ in meters, c is the speed of light, and d_i represents the pseudorange from the *i*th satellite to the receiver. Write out the basic pseudorange equation.

$$\text{Pseudorange}(i) = \sqrt{(x_i - x_r)^2 + (y_i - y_r)^2 + (z_i - z_r)^2} + c(t_{GPS} - t_r)$$

In vector format, let $\vec{D}_i = \vec{R}_i - \vec{R}_r$ be the range vector from the receiver to the *i*th satellite. The pseudorange equation can be written as: $D_i + B = d_i$. Obviously, $R_r^2 = x_r^2 + y_r^2 + z_r^2$.

$$D_i^2 = \vec{D}_i \bullet \vec{D}_i = R_i^2 - 2(x_r x_i + y_r y_i + z_r z_i) + R_r^2 = d_i^2 - 2Bd_i + B^2$$

$$R_i^2 - d_i^2 - B^2 + R_r^2 = 2[x_r x_i + y_r y_i + z_r z_i - Bd_i]$$

Now, subtract between equations for satellites i = 1, 2, 3, 4 or 5 to get:

$$R_2^2 - R_1^2 - (d_2^2 - d_1^2) = 2[(x_2 - x_1)x_r + (y_2 - y_1)y_r + (z_2 - z_1)z_r - (d_2 - d_1)B]$$
$$R_3^2 - R_1^2 - (d_3^2 - d_1^2) = 2[(x_3 - x_1)x_r + (y_3 - y_1)y_r + (z_3 - z_1)z_r - (d_3 - d_1)B]$$
$$R_4^2 - R_1^2 - (d_4^2 - d_1^2) = 2[(x_4 - x_1)x_r + (y_4 - y_1)y_r + (z_4 - z_1)z_r - (d_4 - d_1)B]$$
$$R_5^2 - R_1^2 - (d_5^2 - d_1^2) = 2[(x_5 - x_1)x_r + (y_5 - y_1)y_r + (z_5 - z_1)z_r - (d_5 - d_1)B]$$

Set these 4 difference equations into matrix form and then substitute with matrix symbols.

$$\begin{bmatrix} R_2^2 - R_1^2 - (d_2^2 - d_1^2) \\ R_3^2 - R_1^2 - (d_3^2 - d_1^2) \\ R_4^2 - R_1^2 - (d_4^2 - d_1^2) \\ R_5^2 - R_1^2 - (d_5^2 - d_1^2) \end{bmatrix} = 2 \begin{bmatrix} (x_2 - x_1) & (y_2 - y_1) & (z_2 - z_1) & (d_2 - d_1) \\ (x_3 - x_1) & (y_3 - y_1) & (z_3 - z_1) & (d_3 - d_1) \\ (x_4 - x_1) & (y_4 - y_1) & (z_4 - z_1) & (d_4 - d_1) \\ (x_5 - x_1) & (y_5 - y_1) & (z_5 - z_1) & (d_5 - d_1) \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ B \end{bmatrix} \quad \text{or}$$

$\Re = 2\,\mathbf{H}\,\mathbf{X}$, which is an exact linear matrix equation. If **H** is well conditioned and rank of 4, then $\mathbf{H}^{-1}$ exists, and the solution for $\mathbf{X} = \mathbf{H}^{-1}\Re/2$ is solved immediately. If there are n satellites visible ≥ 5, then the overdetermined matrix equation is solved using least squares. As long as the H matrix is nontrivial (or ill-conditioned) and of rank 4, the matrix product $\mathbf{H}^T\mathbf{H}$ is positive definite and is a symmetric 4 x 4 matrix, which guarantees the inverse. The best estimator using least squares for the unknown vector **X** is:

$$\mathbf{X} = \frac{1}{2}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\,\Re$$

Note that the individual components R_i, x_i, y_i and z_i are obtained directly from the broadcast GPS message, and the d_i, are the measured pseudoranges. With n satellites in view, the $\Re$ is an (n-1) x 1 matrix and **H** is an (n-1) x 4 matrix, where both are obtained by combining both measurements and broadcast values. **X** is the solved matrix, which is 4 x 1.

7.1 Premature Convergence with Standard GPS Iterative Method

The standard GPS Iterative Method usually assumes that the nominal position begins at the center of the Earth and that the solution will converge after a few iterations of the method. This is often done for GPS receivers when no prior knowledge of an approximate position is known for input. This method has been used consistently for over 25 years without any known defects. However, some field tests using ground GPS

transmitters (i.e. pseudolites) showed the GPS receiver converged incorrectly under particular configurations.

7.2 Simulations for Both Methods without Noise in Pseudoranges

The first phase of this analysis examined what empirical characteristics may cause problems with the calculated GPS navigation solutions using the standard GPS method. Six GPS satellites were placed at the corners of a cube where the Earth was located in the cube's center. The distance between the Earth's center and each satellite was at the ideal GPS orbital radius of 26,559,800 meters (GPS JPO, 2000). The receiver was located anywhere near the GPS satellites, and the initial GPS position starting the iterative process was at the Earth's center. These tests assumed no ranging errors from any source, including the atmosphere or clock errors between the GPS satellites and the receiver. Six GPS satellites are located at the vertices of a cube with the edge of the cube being 30668615.36 meters (hereafter, one unit) and the six satellites located at (1,1,1), (1,1,0), (0,1,1), (1,0,1), (1,0,0) and (0,1,0) as shown in Figure 8: Setup of 6 GPS Satellites in Simulation. The Earth's center is located at (0.5, 0.5, 0.5), and the diagonal from the Earth's center to any satellite is 26,559,800 meters. Simultaneously, the exact method ran with the same identical pseudoranges to obtain a navigation solution. The final results from both methods were differenced from the simulated GPS location.

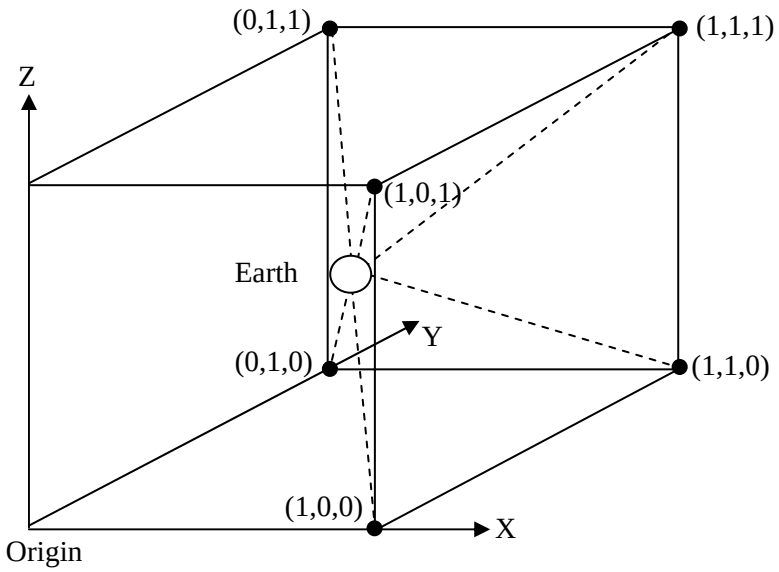

Figure 8: Setup of 6 GPS Satellites in Simulation

Two zones were found along the diagonal between the Earth to the first satellite where the standard GPS navigation solution would not converge within 10 iterations. In the centers, these regions converged to navigation solutions that exceeded thousands of meters from the actual simulated location. The first region was between 5 and 1066 meters away from the satellite at (1,1,1) along the diagonal to the center of the cube with some changing limits. The second region was larger and varied between 231,500 and 2,673,600 meters away from the satellite at (1,1,1) toward the center of the cube as shown.

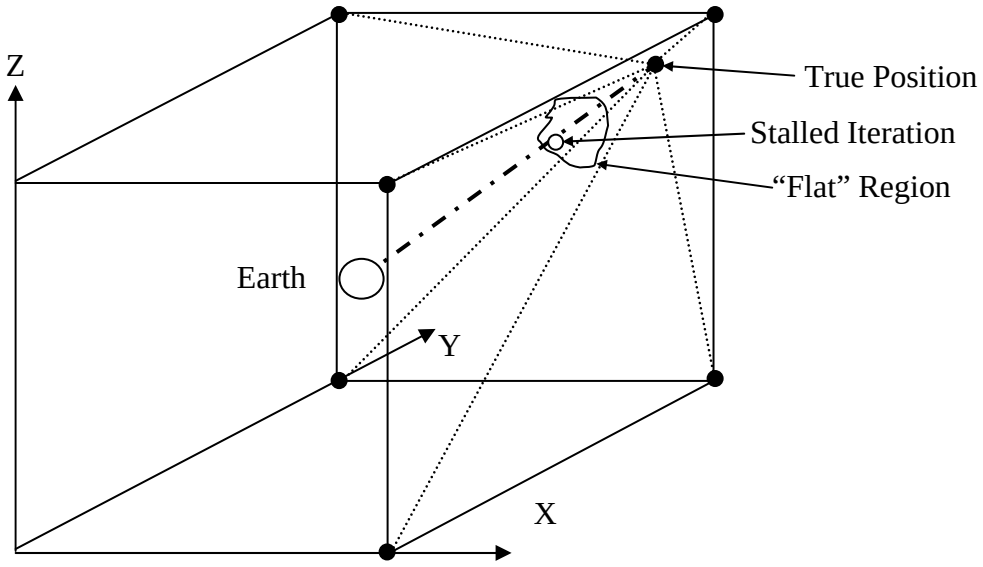

Figure 9: Diagram of Stalled Iteration in "Flat" Region

Separate listings of the empirical limits for 10 and 20 iterations for both zones are shown in Table 2 and Table 3. An example of a stalled iteration is listed in Table 4. Similar zones were found near the second and fourth satellites at approximately the same relative locations where the standard GPS method would iterate suddenly to almost imperceptible increments. The exact method produced a navigation solution within a micron of each test position when assuming no noise in the pseudoranges.

	10 iterations	
X	Y	Z
0.999999	0.999999	0.999996
0.999994	0.999994	0.999978
0.999981	0.999985044	0.999943
0.999980435	0.999980435	0.999941
0.9999999	0.9999999	0.9999997
0.999998	0.999998	0.999993
0.999995	0.999995	0.999982
0.999992	0.999992	0.99997
0.999991	0.999991	0.999965
0.99999	0.99999	0.99997
0.99999	0.99999	0.99996
0.999985	0.999985	0.99997
0.999985	0.999985	0.99996
0.999985	0.999985	0.999945
0.9999825	0.9999825	0.99998
0.9999825	0.9999825	0.99996
0.9999825	0.9999825	0.999943
0.9999815	0.9999815	0.99998
0.9999815	0.9999815	0.99996
0.9999815	0.9999815	0.999942
0.999981	0.999981	0.9999413
0.999981	0.999981	0.99996
0.9999806	0.9999806	0.99997
0.9999806	0.9999806	0.99995
0.9999806	0.9999806	0.999941
0.99998051	0.99998051	0.999941
0.99998051	0.99998051	0.99995
0.999985051	0.999985051	0.99995
0.999985051	0.999985051	0.99995
0.999985051	0.999985051	0.999943
0.999985051	0.999985051	0.999943
0.999985044	0.999985044	0.99995
0.999985044	0.999985044	0.999943
0.999981	0.999985044	0.999943
0.99998044	0.99998044	0.99996
0.99998044	0.99998044	0.999941
0.999980435	0.999980435	0.999941

	20 Iterations	
X	Y	Z
0.999999	0.999999	0.99999
0.999999	0.9999987	0.99999
0.99999	0.99999	0.99999
0.999988	0.9999803	0.9999636
0.999978	0.999978	0.999949

Table 2: Smaller Region Near (1,1,1) Exhibiting Premature Convergence

	10 Iterations	
X	Y	Z
0.99	0.9985	0.998
0.9984	0.99	0.998
0.9984	0.99	0.9907
0.99	0.9975	0.9998
0.96	0.96	0.964245
0.99	0.9975	0.9998
0.99	0.998	0.998
0.92	0.92	0.99
0.94	0.94	0.98
0.95	0.95	0.97
0.96	0.96	0.974245
0.995	0.99671	0.995
0.9967197	0.995	0.995
0.996	0.996	0.995606
0.997	0.994	0.9925

	20 iterations	
X	Y	Z
0.9967196	0.995	0.995
0.995	0.996719603	0.995
0.995	0.995	0.997903772
0.996	0.996	0.9956053
0.996	0.994	0.9905
0.997	0.993	0.9925

Table 3: Larger Region Near (1,1,1) Exhibiting Premature Convergence

Receiver at	(.995,.995,.995)	Initial position at	(0.5,0.5,0.5)
1st Iteration	3390857.7	3390857.7	848913.86
2nd Iteration	252219.07	252219.074	226761.06
3rd Iteration	582810.02	582810.022	449278.59
4th Iteration	586830.34	586830.343	447217.58
5th Iteration	586502.79	586502.785	445771.14
6th Iteration	586413.06	586413.061	445395.72
7th Iteration	586389.80	586389.802	445298.88
8th Iteration	586383.79	586383.794	445273.89
9th Iteration	586382.24	586382.244	445267.44
10th Iteration	586381.84	586381.844	445265.78
11th Iteration	586381.74	586381.741	445265.35
12th Iteration	586381.71	586381.714	445265.24
13th Iteration	586381.71	586381.707	445265.21
14th Iteration	586381.71	586381.705	445265.20
15th Iteration	586381.70	586381.705	445265.20
16th Iteration	586381.70	586381.705	445265.20
17th Iteration	586381.70	586381.705	445265.20
18th Iteration	586381.70	586381.705	445265.20
19th Iteration	586381.70	586381.705	445265.20
20th Iteration	586381.70	586381.705	445265.20
Time Bias	(m) 20^{th} =	-584382.52	

Table 4: Example of Premature Convergence

The exact GPS linear solution worked as expected, except in the region nearby the center of the cube for the center of the Earth or along rays where one component was exactly 0.5 units. These locations prevent the inverse of the $\mathbf{H}^T\mathbf{H}$ matrix, which is ill-conditioned or has 2 identical rows. However, a GPS receiver at the center of the Earth or having the GPS satellites in precise symmetric positions is not realizable in practice, and this does not invalidate the exact method. These same positions would be invalid for the iterative method.

The result of the simulation was the discovery of regions where the GPS navigation iteration converges to positions without getting the receiver's location. The numerical inverse Jacobian is nearly zero approaching these zones, so that the iteration would barely change at the wrong values. Other stalling regions have been found in this simulation, and more are likely elsewhere. These regions are missed when a different starting point is chosen, such as (0.6, 0.7, 0.6), so that the direct line between the start and final true location does not penetrate into any region of early convergence.

Newton's iterative method to solve a set of linearized equations will correctly have both the first and second derivatives at zero when the iteration obtains the actual solution or position of the GPS receiver. However, there is no guarantee in Newton's method that the inverse Jacobian will not be zero from some numerical combination of the partials, which does not require the more stringent condition that the numerical first and second derivatives are nearly zero. Obviously, one can test the solution against the measured pseudoranges to check whether it converged to the correct position or to an unexpected convergence point. Upon detecting a wrong solution, the iteration can be retried with a different starting point to obtain a different converged solution.

These regions of premature convergence are probably the cause of an incorrect GPS solution. The concern is that pseudolites, which are GPS transmitters placed on the ground for improved GPS coverage, may be a detriment rather than an aid for GPS navigation, especially if the FAA is going to place pseudolites at airports for improved GPS precision approach and landings. This study found many regions of early convergence near the theoretical satellite locations, which may indicate that pseudolites aggravate this problem.

Research into the regions of early convergence for the standard GPS iterative method would be very helpful to the GPS community to avoid these convergence zones, but that problem is beyond the scope of this effort. What has been demonstrated is the exact method succeeded in all locations of computing a GPS navigation solution within a micron of the true position in the absence of all noise sources. With the same GPS constellation used in the iterative method analysis, no problems other than the ill-conditioned locations were observed in calculating an exact linear navigation solution at any location. The exact method worked even when the GPS receiver was within the boundaries listed in Tables 1 and 2.

The other facet of the simulation analysis is for GPS receivers in deep space. One of the options being considered is installing a GPS receiver onboard a deep space satellite with additional GPS coverage in deep space. Currently, spaceborne GPS receivers are on satellites in low Earth orbits while the GPS satellites are above the spacecraft. Since the GPS satellites broadcast the signals toward Earth, only GPS satellites that are on the opposite side of the Earth and near its limb do broadcast signals toward GPS receivers that are at higher altitude than the usual GPS orbit. It was discovered that, at sufficient distances away from the Earth (about 2 to 3 times the GPS orbital radius), the GPS navigation solution jumped from a high altitude to a low altitude solution. The standard GPS method in this case uses equations that are approximately linear between the measured ranges and the receiver's position. If the extremely long pseudoranges are almost parallel, then the standard method can not correctly distinguish what are the actual distances in the pseudoranges and what is a clock bias error. The long pseudoranges are shortened when the clock bias is removed, leaving much shorter ranges that make a lower altitude navigation solution. This is due to ignoring the higher order terms in the series expansion of the square root expressions. This example demonstrates that deep space navigation can not use the standard iterative GPS method for a solution, and the exact method is more robust and more accurate for such deep space navigation.

7.3 Simulations for Two Methods with Half-Meter Pseudorange Noise

The simulations inserted a positive half meter noise into the pseudoranges from 3 satellites and the other 3 satellites with a negative half-meter noise. The exact method

was significantly farther from the test location than the standard method by 10 to 100 meters.

It should be pointed out that the exact formulation is more sensitive to any offsets that have not been removed from the measured pseudoranges d_i than the standard GPS iterative method. An inspection of the $\mathfrak{R}$ matrix shows that some components are squares of the pseudorange, which each is between 3.16 and 4.04 times the Earth's radius. Thus, the $\mathfrak{R}$ matrix is susceptible to magnify pseudorange offsets in the navigation solution if the offsets are not removed first. A crude theoretical estimate is the $\mathfrak{R}$ matrix would be affected in each squared d term by 2 d Δd, where Δd is the offset to d in the squared terms of the $\mathfrak{R}$ matrix, and internal differencing can make the Δd terms double, which would made the largest offsets be 4 d Δd. The **H** matrix would have d + Δd terms, so $\mathbf{H}^t$ times $\mathfrak{R}$ should be $8d^2\ \Delta d + 8d\ \Delta d^2$, while the $\mathbf{H}^t\mathbf{H}$ should have 2 d Δd terms.

If the measurement errors can not be reduced or removed, then one optimal solution is to use the exact method to get a close, starting position for the iterative method to obtain the final converged navigation solution. A cross-check in ranges can validate whether the iterative method encountered an unexpected convergence region in its solution.

7.4 Comparison to Other Navigation Solutions for GPS

Stephen Bancroft (1985) published the first paper of an exact algebraic and noniterative method to solve the GPS navigation problem. The derivation reduced the matrix equation of n satellite equations and conditions to a simple scalar quadratic equation, which could be solved by the standard quadratic solution directly. The resulting two scalars would be inserted into the column vector equation to construct the two four-dimensional solutions of position and clock bias for the GPS matrix equation. To distinguish between the two solutions, substitute back into the equations to determine which solution would agree in the matrix equation. Only one of the two solutions would work, and the extraneous solution would be discarded. The matrix solution relied on the least squares formulation, and the method was efficient. However, it required the extra step of back substitution to eliminate the extraneous result, and it was not a linear

formulation that could be applied to most Kalman filters that are predominately used in smoothing navigational data.

Lloyd Krause (1987) developed an iterative technique with difference equations that were linearized. In the development, a quadratic equation for the nominal clock bias was given, but only the root with the negative square root was used. No physical reason was given why the positive root was discarded. In reality, the receiver clock bias can be either positive or negative with even large absolute magnitudes. As the technique was iterative, it probably relied on the convergence capability by iteration to obtain the correct final answer.

Biton, Koifman and Bar-Itzhack (1996) published an iterative solution that was similar to Bancroft's solution with the difference of computing a different pseudoinverse matrix. The paper claimed a convergence within 2 to 3 iterations, while the standard GPS least squares method took 5 to 6 iterations to converge. These authors showed Bancroft's method was computationally deviated about 5 meters from the standard GPS method, while their iterative formulation was within half a meter of the standard GPS method. The analysis did not compare any method to a surveyed benchmark for accuracy. So, the method is probably a bit faster than the standard GPS method, but it is not so different to warrant its implementation in GPS receiver designs.

Chapter 8 Precise Pseudoranges with Dual Phase and Code

The analysis of the exact method showed sensitivity to measurement errors that drove the navigation solution further from truth than the iterative method, especially when the pseudorange errors exceeded a meter. So, this analysis shows a method to obtain precise pseudorange measurements with minimal errors by combining standard code measurements with the dual carrier phase measurements. Although it is not possible to reconstruct the precise input GPS signal directly after dispreading, the tracking loops for the code and carrier do generate the expected input signal characteristics, which are used for correlation against the input signals. As pointed out before, the internal phase angle is the arctangent of the quadrature divided by the in-phase signal. Knowing that the GPS signal has the code demarked by 180 degree phase shifts in the carrier phase signals, the carrier phase angle of the carrier tracking loop may be any arbitrary angle at the moment of the phase discontinuity of 180 degrees. However, this demarcation is consistently repeated at subsequent code boundary occurrences over the short term. Subtle changes in the atmosphere and the rotation of the GPS satellite to keep both the Sun and Earth in view will rotate the phase angle without doing any physical changes to the GPS signal.

8.1 Enhanced Pseudorange Measurement using Carrier Phases

The goal of this phase of the research is to provide enhanced pseudorange measurements to centimeter precision using dual carrier phase angle measurements. The current way of using carrier phase measurements is through DGPS (usually postprocessed) with two GPS receivers, one base station usually stationary at a known benchmark and the other a roving remote collecting data simultaneously. A new method documented in this thesis uses only one GPS receiver and performs the final calculations in near real time.

8.2 Derivation of Range Adjustment using Dual Carrier Phases

The GPS design specifically chose the frequencies of L1 and L2 to be such that $f_{L1}/f_{L2} = 77/60$ exactly. Another way of looking at this is by wavelengths, λ, where:

$$77\lambda_{L1} = 60\lambda_{L2} \approx 14.65^{+} \text{ meters}$$

Use the pseudorange measurement after removing the clock bias to get a reasonably good estimate of the range between the GPS satellite and the receiver. After removing all known range errors, the remaining range error is usually much less than 3 meters (below a 1/10 of a P(Y) chip), so there is plenty of margin to determine how many whole intervals of 14.65 meters are possible between the satellite and the receiver. Assume for now that the phase angles of L1 and L2 are Φ'_{L1} and Φ'_{L2} precisely after a boundary shift of the P(Y) chip. This phase angle is usually nonzero because the GPS satellite slowly rotates once per orbit.

Calculate K = Truncate[Distance/14.65] with the distance in meters. Define the following terms for the derivation:

M is an integer where M = 77K so that $M\lambda_{L1} = 77K\lambda_{L1} = 14.65K \le \text{Distance}$

N is an integer where N = 60K so that $N\lambda_{L2} = 60K\lambda_{L2} = 14.65K \le \text{Distance}$

B is the time bias in meters = c x time offset between the satellite and receiver clocks.

m and n are integers such that $0 \le m < 77$ and $0 \le n < 60$

ϕ_{L1} and ϕ_{L2} are part of the phase measurement and less than a revolution.

It is noted here that the phase measurements are corrected for the phase angle offset through the ionosphere before doing this calculation. The total electron content is obtained with dual frequency measurements. By multiplying with the proper coefficient, the proper correction to the measured phase ϕ' is obtained. Any nonzero phase for Φ' will cause all ϕ' data to be corresponding larger than expected. Convert the phase measurements by $\phi_{L1} = \phi'_{L1} - \Phi'_{L1}$.

$$(M + m + \phi_{L1})\lambda_{L1} + B = (N + n + \phi_{L2})\lambda_{L2} + B$$

$(M + m)\lambda_{L1} - (N + n)\lambda_{L2} = \phi_{L2}\lambda_{L2} - \phi_{L1}\lambda_{L1}$ and divide by λ_{L1} to get

$$M + m - \frac{77}{60}N = \frac{77}{60}n + \frac{77}{60}\phi_{L2} - \phi_{L1}$$

Similarly, $$N + n - \frac{60}{77}M = \frac{60}{77}m + \frac{60}{77}\phi_{L1} - \phi_{L2}$$

Now the quantity on the left side of the equation is an integer, so the quantity on the right is also an integer. So, try all possible values of n as $0 \le n < 60$ and find the one value of n

that makes the quantity $\frac{77}{60}n + \frac{77}{60}\phi_{L2} - \phi_{L1}$ the closest to an integer. Use the other equation for determining m as a consistency check. Then, within measurement error, the solution is:

Distance $D_i = 14.65\ K + (n + \phi_{L2})\lambda_{L2}. = 14.65\ K + (m + \phi_{L1})\lambda_{L1}$ in meters.

The error in the above equation is determined by the error in the phase measurement, which should be no worse than 3 degrees in ϕ or, equivalently, less than 3 mm in wavelength. With each L1 revolution in phase, L2 will advance 60/77 of a revolution. Assume the L1 and L2 phases are zero at the code boundary. Figure 10 shows the 360° revolutions of 0 to 76 in L1 through a half chip. Figure 11 shows where L2 is at each of the n revolutions of L1. For any L2 location at integer L1 revolutions, the neighboring L2 locations in Figure 11 are 9 revolutions from the original L2 location and are 360°/77 = 4.675° apart.

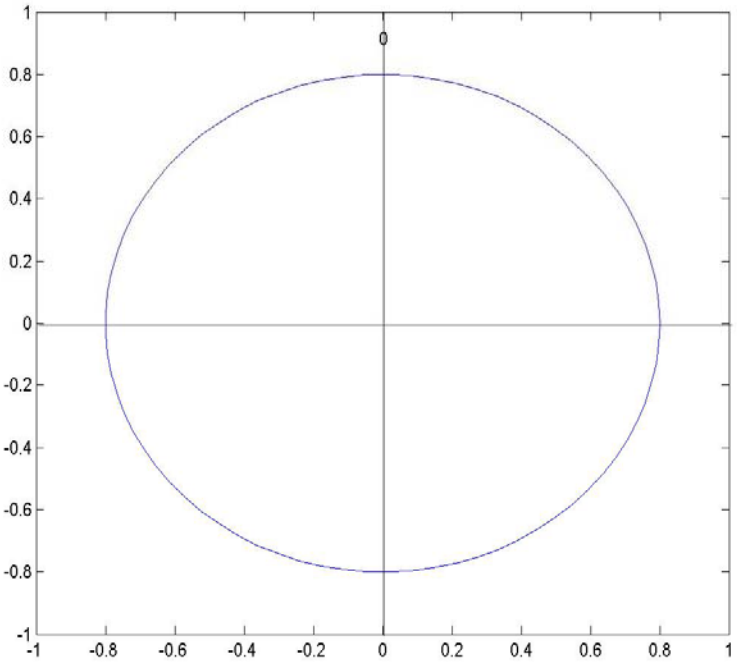

Figure 10: Phase Diagram of L1 Whole Revolutions Starting at Zero

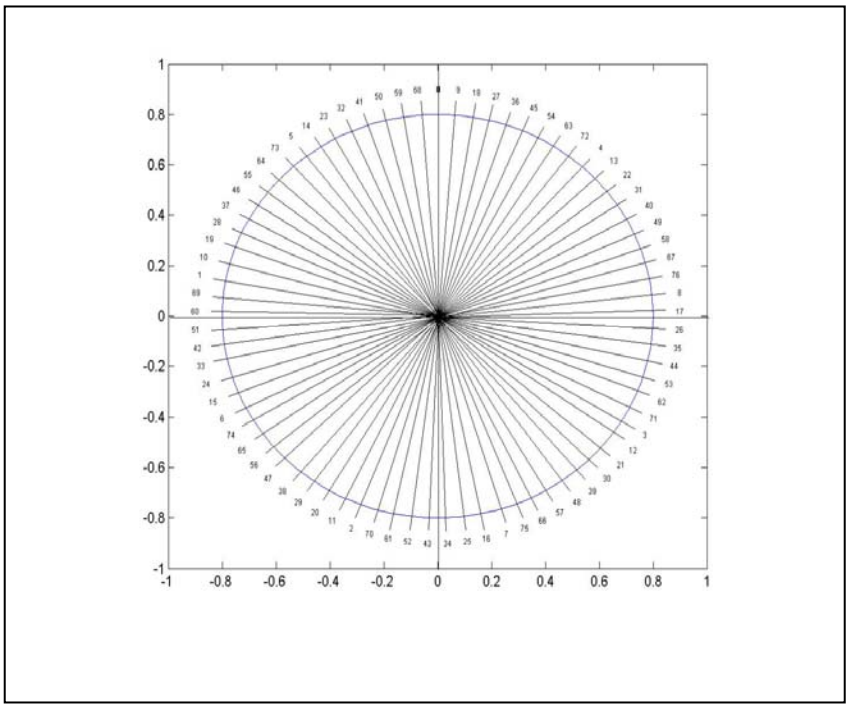

Figure 11: L2 Phase Diagram Relative to Numbered Whole L1 Revolutions

If both L1 and L2 phase angles are measured within an uncertainty of 2.3°, then each combination of L1 and L2 phase angles is unique within the half chip interval to an uncertainty of 3 mm. However, suppose the phase noise is large enough so that at a normal 99% confidence interval designated by a 3σ statistic exceeds ±4 degrees, then two phase possibilities may occur. In that case, the separation of 9 revolutions of L1 means there is a distance of 9 x 19 cm = 1.7126 m in wavelength between possibilities. Referring to the precise pseudorange measurements from the code tracking loop to an accuracy less than 0.85 meters would guarantee the determination of which pair of phase angles is the correct one.

The final check is made that the initial pseudorange distance after corrections ≈ Distance D_i, since the truncation could cause a round off of one span. For example, if the initial distance was 14.65(k + 0.975) with an actual pseudorange error of 0.05 of a half chip (i.e. 0.85 m error), then the truncation would get k intervals instead of k+1 intervals. In that case, one final adjustment in the value of K is warranted in the distances.

8.3 Experiment Using Pseudorange Adjustment and Phases

Actual differential GPS (DGPS) receiver tests at known benchmarks are chosen to evaluate this theoretical method of adjusting pseudorange measurements with dual phase data. Ashtech GPS survey receivers provided the DGPS data that were made on March 19, 2002, at 3 locations inside the city limits of Ames, Iowa. These data are analyzed in detail in the next chapter. These particular receivers do not directly measure the military P(Y) code simultaneously on L1 and L2, but they rely on a codeless technique of squaring the input P(Y) signal with itself to determine where the pseudoranges are within the C/A chip. P(Y) code chips are 1/10 of the size of the C/A code chip, and navigation solutions using P(Y) code should be more precise. Manual processing of the actual GPS measurements were corrected for ionosphere, troposphere, relativity and Earth rotation. The navigation solutions were compared using the proposed techniques and the standard GPS method against the survey benchmarks identified as IADO, G506 and G001. IADO is the Iowa Department of Transportation building located approximately in the center of the Ames, G506 is located in the north of the Iowa State University (ISU) central campus lawn, and G001 is about 10 meters north of the Town Engineering Building of ISU. G506 is 1.984 km west and slightly north of IADO, and G001 is 672 m northwest of G506 on the ISU campus.

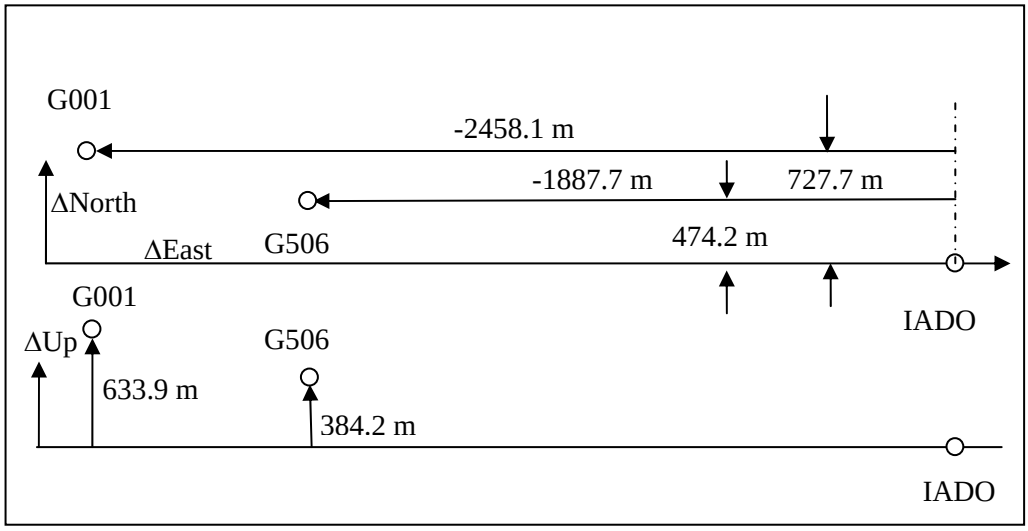

Figure 12: Three Benchmarks in Ames in Relative Local Coordinates

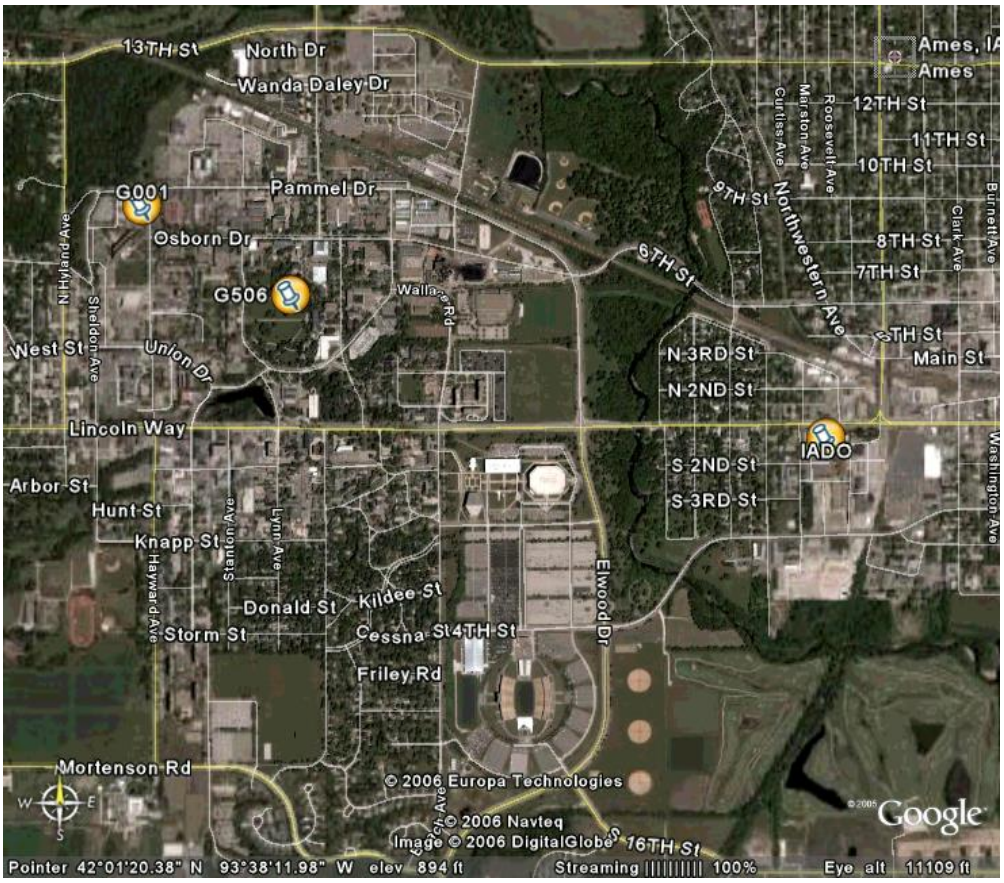

Figure 11: Arial Map Showing Benchmarks in Ames

SV #	Azimuth	Elevation
2	80.79	23.66
8	-57.31	68.73
11	-48.71	26.36
27	-29.26	74.94
28	-83.38	40.84
31	59.34	49.36

Table 5: Local Satellite Coverage at Ames

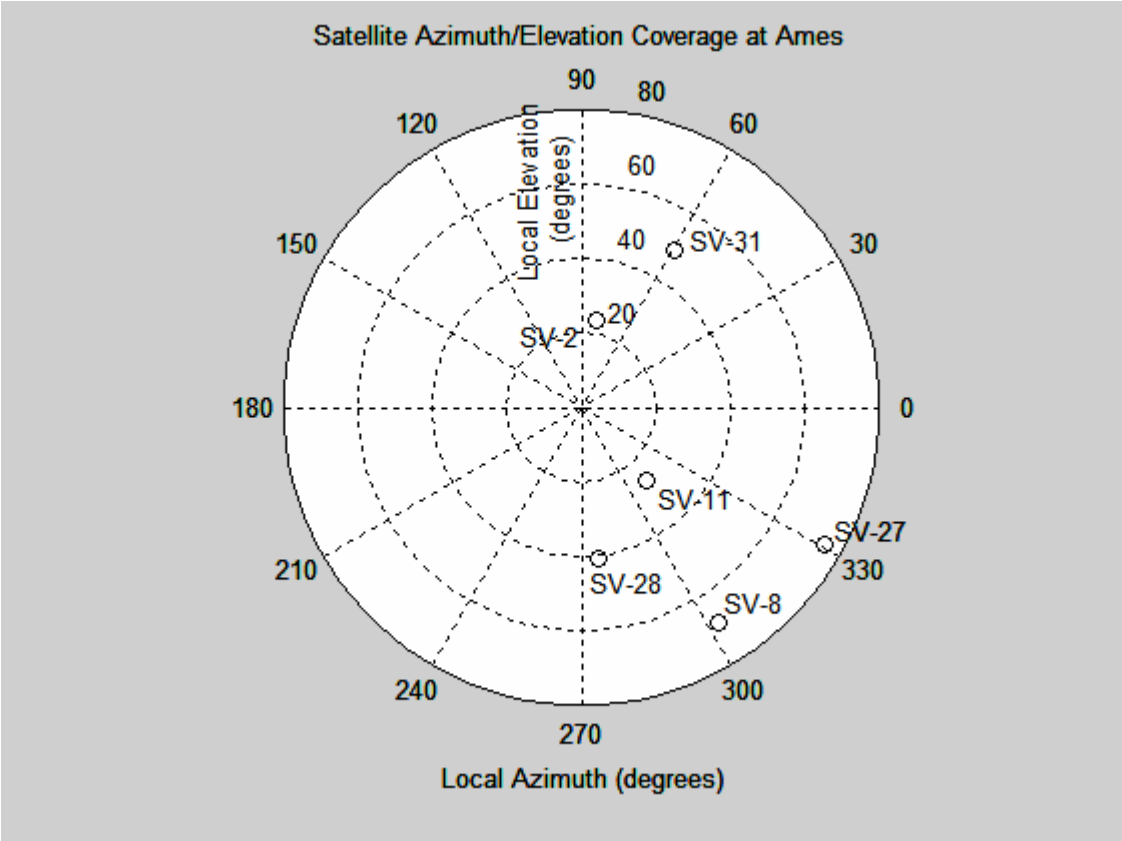

Figure 13: GPS Satellite Constellation during Survey

Chapter 9 Analysis of GPS Survey Data

Data from several Ashtech receivers were compiled over several hours for DGPS surveys to determine the precise altitude of various points within Story County in six month intervals for about three years. Data from three Ashtech receivers were obtained for this study at the Department of Transportation Headquarters Building (designated IADO), a point on the central campus of Iowa State University (designated G506), and a point just about ten meters north of Town Engineering Building of Iowa State University (designated G001). Because Town Engineering Building blocks satellites south of G001, the maximum number of common satellites was 6 with only 90 seconds of continuous data collected at all three sites simultaneously. This greatly affected the navigation solutions to vary significantly from their benchmarks over the short period of data recording. The data are collected every fifteen seconds and are listed in Table 6 with the Jet Propulsion Laboratory (JPL) postprocessed GPS satellite ephemeredes spaced at fifteen minute intervals, which JPL claims are accurate to three centimeters. The satellite ephemeredes were obtained by Lagrangian interpolation at both the time of reception and then iterated backwards to the time of transmission (refer to Appendix). Pseudorange data are listed as C/A L1 and P(Y) L2 from all three locations beginning at 48700 s through 48855 s of that day from satellites PRN 2, 8, 11, 27, 28 and 31. Carrier phase data are also listed under the columns labeled as L1 Car. and L2 Car.

Selected Data for Analysis			19-Mar-02		JPL Ephmeredes (Position)			
G506	L1 Car.	C/A L1	L2 Car.	P L2	ECEF X	ECEF Y	ECEF Z	Sat Clock
48780	13 h	33 m	0 s				Epoch	46800
2	4259791	23663850	3315611	23663860	16804.046	-11050.58	17800.123	-169.50377
8	-2234620	20263631	-1736808	20263637	-10693.787	-10702.765	21603.509	707.90108
11	-3805609	23206234	-2959429	23206244	13045.369	-22523.615	-5242.1493	9.826568
27	845850.9	20342736	652590.5	20342741	-1612.8446	-19031.802	18533.267	14.096344
28	-3018686	22252898	-2347274	22252903	-20789.105	-12588.224	11107.646	-71.400002
31	3838096	21294376	2988240	21294382	8593.0273	-16351.979	18685.633	100.42612
48795	13 h	33 m	15 s				Epoch Sec.	47700
2	4301629	23671811	3348212	23671823	18619.834	-10651.359	16047.97	-169.5092
8	-2247903	20261104	-1747159	20261110	-8982.4236	-12537.893	21378.484	707.89771
11	-3836406	23200374	-2983426	23200383	13610.756	-22659.577	-2520.2196	9.826712
27	862242.6	20345855	665363.2	20345861	-229.72613	-20483.066	16903.636	14.097469
28	-3041517	22248554	-2365064	22248559	-19396.004	-12525.722	13453.411	-71.402405
31	3881357	21302608	3021950	21302614	9565.9748	-14289.186	19879.036	100.42878

Table 6: Pseudorange Data from Three GPS Survey Receivers

Selected Data for Analysis			19-Mar-02		JPL Ephemerides (Position)			
G506	L1 Car.	C/A L1	L2 Car.	P L2	ECEF X	ECEF Y	ECEF Z	Sat Clock
48810	13 h	33 m	30 s				Epoch Sec.	48600
2	4343503	23679779	3380841	23679793	20249.344	-10345.57	14026.703	-169.51462
11	-3867126	23194529	-3007364	23194537	13997.252	-22564.775	245.12365	9.827224
8	-2261065	20258599	-1757415	20258605	-7395.9409	-14422.785	20777.246	707.89433
27	878770.8	20349001	678242.3	20349007	971.79551	-21839.301	14983.082	14.098595
28	-3064294	22244219	-2382812	22244225	-17752.693	-12504.743	15571.826	-71.404656
31	3924736	21310863	3055751	21310869	10684.814	-12171.34	20719.525	100.43093
48825	13 h	33 m	45 s				Epoch Sec.	49500
2	4385411	23687755	3413496	23687767	21658.76	-10108.91	11768.31	-169.52004
8	-2274104	20256118	-1767575	20256124	-5959.0056	-16314.021	19809.968	707.89096
11	-3897771	23188697	-3031243	23188704	14231.577	-22218.752	3006.2463	9.827736
27	895435.2	20352171	691227.5	20352178	1991.4491	-23056.7	12803.037	14.09972
28	-3087018	22239895	-2400520	22239901	-15887.378	-12556.555	17427.053	-71.406907
31	3968233	21319140	3089645	21319146	11938.905	-10045.798	21192.776	100.43308
48840	13 h	34 m	0 s				Epoch Sec.	50400
2	4427351	23695736	3446177	23695748	22819.992	-9912.1286	9309.2865	-169.52904
8	-2287025	20253659	-1777643	20253665	-4688.5021	-18165.408	18493.44	707.88758
11	-3928343	23182879	-3055065	23182887	14344.655	-21608.097	5715.6058	9.828247
27	912232.8	20355368	704316.5	20355375	2836.7205	-24093.48	10399.881	14.116831
28	-3109693	22235580	-2418188	22235586	-13834.769	-12707.311	18987.605	-71.409158
31	4011844	21327439	3123627	21327445	13309.866	-7958.0838	21291.201	100.43522
48855	13 h	34 m	15 s					
2	4469322	23703723	3478881	23703733				
8	-2299827	20251223	-1787619	20251229				
11	-3958841	23177074	-3078830	23177084				
27	929163.3	20358589	717509.1	20358596				
28	-3132318	22231276	-2435818	22231280				
31	4055568	21335760	3157698	21335766				

Selected Data for Analysis			19-Mar-02		JPL Ephemerides (Velocity)			
IADO	L1 Car.	C/A L1	L2 Car.	P L2	ECEF X Vel.	ECEF Y Vel.	ECEF Z Vel.	
48780	13 h	33 m	0 s				Epoch Sec.	46800
2	930492	23412072	674956.7	23412083	21071.854	5028.9044	-17868.151	
8	-4374265	20014193	-3388455	20014198	19597.703	-19964.331	-391.33646	
11	-5538788	22954497	-4293776	22954507	7360.4451	-2699.2566	29827.63	
27	-1779716	20092244	-1375555	20092251	16353.978	-16488.968	-16389.099	
28	-4988397	22004250	-3858792	22004256	14004.55	797.96589	27183.565	
31	673627	21043379	523720.9	21043386	9982.9452	22437.424	15149.362	
48795	13 h	33 m	15 s				Epoch Sec.	47700
2	941439.2	23414156	683486.9	23414167	19207.805	3876.8735	-21018.299	
8	-4418446	20005786	-3422882	20005791	18373.844	-20743.702	-4602.9431	
11	-5600465	22942760	-4341836	22942771	5243.3561	-272.62265	30572.685	
27	-1794219	20089484	-1386856	20089491	14366.904	-15679.283	-19776.614	
28	-5042116	21994028	-3900651	21994034	16914.619	525.73918	24871.085	
31	686004.4	21045734	533365.6	21045741	11633.327	23314.212	11332.059	

Table 6: Pseudorange Data from Three GPS Survey Receivers (Continued)

Selected Data for Analysis			19-Mar-02		JPL Ephmeredes (Velocity)			
IADO	L1 Car.	C/A L1	L2 Car.	P L2	ECEF X Vel.	ECEF Y Vel.	ECEF Z Vel.	
48810	13 h	33 m	30 s				Epoch Sec	48600
2	952405.8	23416243	692032.1	23416253	16940.663	2963.611	-23839.45	
8	-4462520	19997399	-3457225	19997404	16835.49	-21062.082	-8739.0314	
11	-5662082	22931035	-4389849	22931045	3395.1488	2417.3787	30791.017	
27	-1808602	20086747	-1398064	20086754	12333.088	-14377.83	-22844.072	
28	-5095798	21983813	-3942481	21983819	19551.012	-117.67561	22138.327	
31	698484.3	21048109	543090.1	21048116	13209.995	23661.014	7318.9013	
48825	13 h	33 m	45 s				Epoch Sec	49500
2	963386	23418333	700588	23418343	14327.487	2349.8911	-26279.392	
8	-4506493	19989031	-3491490	19989036	15065.04	-20879.524	-12724.896	
11	-5723644	22919320	-4437820	22919330	1869.5983	5296.9679	30479.064	
27	-1822870	20084032	-1409182	20084039	10339.815	-12597.794	-25533.907	
28	-5149448	21973603	-3984286	21973609	21835.936	-1082.3523	19030.535	
31	711060.6	21050503	552889.8	21050509	14624.546	23488.637	3183.7135	
48840	13 h	34 m	0 s				Epoch Sec	50400
2	974389.9	23420427	709162.4	23420437	11438.028	2084.3779	-28290.388	
8	-4550356	19980684	-3525669	19980690	13152.431	-20174.582	-16488.39	
11	-5785141	22907617	-4485740	22907628	706.02239	8283.8395	29642.421	
27	-1837012	20081341	-1420201	20081348	8471.5958	-10370.897	-27793.895	
28	-5203057	21963402	-4026059	21963408	23703.137	-2304.8346	15599.002	
31	723743.1	21052916	562772.3	21052923	15793.844	22826.88	-997.88452	
48855	13 h	34 m	15 s					
2	985411	23422524	717750.1	23422535				
8	-4594114	19972357	-3559766	19972363				
11	-5846579	22895926	-4533613	22895936				
27	-1851035	20078673	-1431129	20078679				
28	-5256630	21953207	-4067805	21953213				
31	736525	21055348	572732.1	21055355				

Selected Data for Analysis			19-Mar-02	
G001	L1 Car.	C/A L1	L2 Car.	P L2
48780	13 h	33 m	0 s	
2	244897.5	23504276	180981.2	23504278
8	-325696	20103325	-239216	20103324
11	-521567	23046779	-380417	23046783
27	13701.84	20182827	6063.299	20182827
28	-430960	22092373	-313216	22092372
31	309128.3	21134525	228441	21134526
48795	13 h	33 m	15 s	
2	272986.5	23509621	202868.7	23509624
8	-352726	20098181	-260279	20098180
11	-566118	23038301	-415132	23038305
27	16346.16	20183331	8123.762	20183330
28	-467542	22085412	-341722	22085411
31	338635.3	21140140	251433.4	21140140

Table 6: Pseudorange Data from Three GPS Survey Receivers (Continued)

Selected Data for Analysis			19-Mar-02	
G001	L1 Car.	C/A L1	L2 Car.	P L2
48810	13 h	33 m	30s	
2	301196.3	23514989	224850.3	23514992
8	-379549	20093077	-281180	20093076
11	-610508	23029853	-449722	23029857
27	19212.57	20183876	10357.29	20183876
28	-503986	22078477	-370119	22078477
31	368346.1	21145794	274584.7	21145795
48825	13 h	33 m	45 s	
2	329519.9	23520378	246920.4	23520382
8	-406171	20088010	-301925	20088011
11	-654743	23021436	-484191	23021440
27	22294.18	20184463	12758.51	20184461
28	-540298	22071567	-398414	22071567
31	398253.6	21151485	297889.1	21151486
48840	13 h	34 m	0 s	
2	357952	23525789	269075.2	23525793
8	-432598	20082982	-322517	20082982
11	-698828	23013047	-518543	23013051
27	25586.08	20185089	15323.59	20185088
28	-576483	22064681	-426611	22064680
31	428352.1	21157213	321342.5	21157214
48855	13 h	34 m	15 s	
2	386487.4	23531220	291310.5	23531223
8	-458833	20077990	-342960	20077990
11	-742768	23004685	-552781	23004690
27	29083.03	20185754	18048.44	20185754
28	-612547	22057818	-454712	22057817
31	458636.2	21162976	344940.5	21162977

Table 6: Pseudorange Data from Three GPS Survey Receivers (Continued)

The pseudorange residuals between the actual distance taken from surveyed benchmarks to the satellites and the corrected pseudoranges were tens of meters off except for PRN-31, as shown in the bottom third of Table 7 through Table 9. The residuals demonstrate that corrections must be applied to the measured pseudoranges to improve navigation accuracy.

Round Res.	Time 48780	Time 48795	Time 48810	Time 48825	Time 48840	Time 48855
PRN 2	-2	-2	-2	-2	-1	-1
PRN 8	3	3	3	3	3	3
PRN 11	-1	-1	-1	-1	-1	-1
PRN 27	1	1	1	1	1	1
PRN 28	3	3	3	3	3	3
PRN 31	0	0	0	0	0	0
Truncate Res.	Time 48780	Time 48795	Time 48810	Time 48825	Time 48840	Time 48855
PRN 2	-1	-1	-1	-1	-1	-1
PRN 8	2	3	3	3	3	3
PRN 11	0	0	0	0	0	0
PRN 27	1	1	1	1	1	1
PRN 28	3	3	3	3	3	3
PRN 31	0	0	0	0	0	0
Residual (m)	Time 48780	Time 48795	Time 48810	Time 48825	Time 48840	Time 48855
PRN 2	-25.55858	-24.71059	-23.44986	-22.16336	-20.94090	-19.78103
PRN 8	43.84553	44.71796	45.72858	46.72312	47.66823	48.62993
PRN 11	-10.73449	-10.07085	-9.450774	-8.888415	-8.306192	-7.697212
PRN 27	15.36069	16.37646	17.68685	18.94695	20.16710	21.36208
PRN 28	49.35315	49.58462	49.93076	50.23911	50.53947	50.8508
PRN 31	1.329365	1.559243	2.28351	2.968408	3.644334	4.300024

Table 7: Half Chip Adjustments with Ashtech Pseudorange Data at IADO

Round Res.	Time 48780	Time 48795	Time 48810	Time 48825	Time 48840	Time 48855
PRN 2	-2	-2	-1	-1	-1	-1
PRN 8	3	3	3	3	3	3
PRN 11	-1	-1	-1	-1	-1	-1
PRN 27	1	1	1	1	1	1
PRN 28	3	3	3	3	3	3
PRN 31	0	0	0	0	0	0
Truncate Res.	Time 48780	Time 48795	Time 48810	Time 48825	Time 48840	Time 48855
PRN 2	-1	-1	-1	-1	-1	-1
PRN 8	2	2	2	3	3	3
PRN 11	0	0	0	0	0	0
PRN 27	0	0	1	1	1	1
PRN 28	3	3	3	3	3	3
PRN 31	0	0	0	0	0	0
Residual (m)	Time 48780	Time 48795	Time 48810	Time 48825	Time 48840	Time 48855
PRN 2	-25.43388	-23.28984	-21.27377	-19.99980	-18.09080	-18.11952
PRN 8	42.61330	43.05491	43.43300	45.0101	47.37709	48.53149
PRN 11	-12.61064	-13.12165	-13.15850	-13.30548	-10.81611	-9.081680
PRN 27	12.96033	14.11648	15.42259	16.50468	19.93002	21.15018
PRN 28	47.61779	47.56238	47.56710	48.18679	49.54814	50.33297
PRN 31	-0.41863	-0.262763	0.334074	1.00563	3.290256	4.763457

Table 8: Half Chip Adjustments with Ashtech Pseudorange Data at G056

Round Res.	Time 8780	Time 48795	Time 48810	Time 48825	Time 48840	Time 48855
PRN 2	-2	-2	-2	-2	-1	-1
PRN 8	3	3	3	3	3	3
PRN 11	-1	-1	-1	-1	-1	-1
PRN 27	1	1	1	1	1	1
PRN 28	3	3	3	3	3	3
PRN 31	0	0	0	0	0	0
Truncate Res.	Time 48780	Time 48795	Time 48810	Time 48825	Time 48840	Time 48855
PRN 2	-1	-1	-1	-1	-1	-1
PRN 8	2	3	3	3	3	3
PRN 11	0	0	0	0	0	0
PRN 27	1	1	1	1	1	1
PRN 28	3	3	3	3	3	3
PRN 31	0	0	0	0	0	0
Residual (m)	Time 48780	Time 48795	Time 48810	Time 48825	Time 48840	Time 48855
PRN 2	-24.62451	-23.79381	-22.99682	-22.25389	-19.68274	-18.60185
PRN 8	43.91640	44.37117	45.00605	45.63302	47.9118	49.39467
PRN 11	-11.4196	-11.22234	-11.55069	-11.47921	-9.194201	-8.246747
PRN 27	15.36476	16.15085	16.61266	16.53064	19.43826	20.83736
PRN 28	49.23460	49.3547	49.57811	49.18919	49.86694	50.34640
PRN 31	2.076359	1.932424	2.13182	2.191678	4.915539	5.526159

Table 9: Half Chip Adjustments with Ashtech Pseudorange Data at G001

The rounding and truncation results are shown in the upper two thirds of Table 7 through Table 9. After standard corrections were applied, the next processing step modified the pseudoranges by adding integer lengths of the half-chip (14.65 m) by rounding or by truncation. This reduced the overall pseudorange residuals so that the solved navigation solutions were closer to the benchmarks. Both approaches were used with all carrier phase angle data in fractional cycles of a revolution. Since there was no means of capturing the actual phase angle at the code boundary, it was assumed to be zero for this analysis. The procedure described in Chapter 9 was used to determine the unique location of the L1 and L2 fractional carrier phase pairing within the half chip. This solution for each receiver at each time point was compared to the associated benchmark to produce the listed XYZ residuals in Table 10 through Table 12. Under the time heading is the total RMS residual broken into XYZ coordinates using the updated pseudoranges.

IADO	Time 48780	Time 48795	Time 48810	Time 48825	Time 48840	Time 48855
Truncation without phase correction						
RMS of ΔXYZ	27.849	29.304	33.519	38.494	39.389	32.275
ΔX	6.651	11.113	2.676	-0.459	1.255	-4.989
ΔY	25.318	25.991	30.39	37.609	24.888	23.698
ΔZ	-9.503	-7.729	-13.886	-8.196	-30.504	-21.335
Truncation with phase correction						
RMS of ΔXYZ	23.159	22.846	32.787	34.11	34.162	25.35
ΔX	0.9314	15.832	5.985	2.377	8.184	-0.102
ΔY	22.42	16.456	30.918	33.972	19.131	20.669
ΔZ	-5.728	-0.684	-9.123	-1.938	-27.094	-14.678
Rounding without phase correction						
RMS of ΔXYZ	18.873	19.223	20.07	28.905	32.609	28.365
ΔX	0.166	6.535	-1.867	-4.967	1.076	-5.155
ΔY	10.372	15.237	19.647	26.876	28.167	26.947
ΔZ	15.766	9.729	3.646	9.41	-16.395	-7.202
Rounding with phase correction						
RMS of ΔXYZ	21.647	20.988	21.905	28.108	27.109	23.9254
ΔX	-5.553	11.253	1.441	-2.13	8.005	-0.269
ΔY	7.475	5.703	20.175	23.239	22.41	23.918
ΔZ	19.542	16.774	8.409	15.668	-12.985	-0.544

Table 10: Phase Corrections Applied to Ashtech Data at IODC

G506	Time 48780	Time 48795	Time 48810	Time 48825	Time 48840	Time 48855
Truncation without phase correction						
RMS of ΔXYZ	44.456	24.155	15.885	16.141	8.798	21.878
ΔX	5.456	9.781	15.463	8.608	5.543	-6.042
ΔY	31.051	17.123	-3.614	-9.824	-5.454	16.766
ΔZ	-31.343	-13.95	-0.404	9.483	4.116	-12.689
Truncation with phase correction						
RMS of ΔXYZ	51.283	36.336	18.282	9.482	11.237	21.632
ΔX	0.594	8.235	14.911	5.973	8.009	-5.968
ΔY	34.697	28.977	9.693	-6.982	-5.893	16.643
ΔZ	-37.759	-20.318	-4.233	2.343	5.234	-12.464
Rounding without phase correction						
RMS of ΔXYZ	6.357	28.7291	25.895	25.86	19.123	21.005
ΔX	0.5166	4.923	13.434	8.417	5.365	-6.207
ΔY	-5.602	-19.542	-4.604	-6.516	-2.175	20.015
ΔZ	2.959	20.475	21.654	23.568	18.226	1.444
Rounding with phase correction						
RMS of ΔXYZ	5.887	16.417	23.652	17.799	21.032	20.8836
ΔX	-4.346	3.377	12.881	5.782	7.831	-6.134
ΔY	-1.956	-7.687	8.703	-3.674	-2.614	19.892
ΔZ	-3.456	14.107	17.825	16.428	19.343	1.67

Table 11: Phase Corrections Applied to Ashtech Data at G506

G001	Time 48780	Time 48795	Time 48810	Time 48825	Time 48840	Time 48855
Truncation without phase corrections						
RMS of ΔXYZ	31.631	19.243	33.822	48.399	13.499	15.187
ΔX	2.203	4.303	4.985	5.142	9.897	-7.723
ΔY	21.934	15.918	33.037	46.536	-1.162	1.541
ΔZ	-22.684	-9.920	-5.254	-12.267	-9.106	-12.986
Truncation with phase corrections						
RMS of ΔXYZ	16.026	17.794	39.916	25.882	19.387	12.006
ΔX	7.441	-3.645	1.399	8.503	11.902	-8.909
ΔY	11.899	4.886	36.725	23.922	13.677	0.269
ΔZ	-7.738	-16.717	-15.577	5.032	-6.8651	-8.045
Rounding without phase corrections						
RMS of ΔXYZ	8.592	9.142	25.4553	36.203	11.134	9.299
ΔX	-4.28	-0.2747	0.443	0.636	9.719	-7.888
ΔY	6.986	5.163	22.293	35.802	2.117	4.789
ΔZ	2.587	7.538	12.278	5.34	5.003	1.147
Rounding with phase corrections						
RMS of ΔXYZ	17.822	10.129	26.243	26.503	21.85	11.48
ΔX	0.958	-8.222	-3.142	3.997	11.724	-9.074
ΔY	-3.049	-5.869	25.981	13.188	16.956	3.517
ΔZ	17.534	0.742	1.956	22.638	7.245	6.089

Table 12: Phase Corrections Applied to Ashtech Data at G001

At all 3 sites, the rounding of integer lengths of half chip intervals and adding that to the original pseudoranges produced less residuals than with the truncation technique. The decision criteria were obtained by summing the total residuals over all six times for the two techniques.

The next processing step now incorporates the carrier phase adjustments as given in Chapter 9 produced the additional average deviations given in Table 13. Phase corrections incorporated with the rounded half chip intervals produced lower navigation residuals for the IODC and G506 data when differenced with the benchmarks, but the opposite is true for G001. Phase measurements incorporated with truncated half chip intervals improved the navigation residuals for IADO and G001, but they made it worse for G506. Incorporating phase measurements improved 2 of the 3 data sites by either rounding or truncation. Thus, the GPS data did not demonstrate a definite improvement with the dual carrier phase method outlined in Chapter 9.

Method	IADO		G506		G001	
	With	Without	With	Without	With	Without
Round	23.947	24.674	17.611	21.161	19.004	16.637
Truncation	28.735	33.471	24.788	21.885	21.835	26.963

Table 13: Average Deviations With and Without Phase measurements

Differential GPS results are tabulated using either method to analyze the data for any further improvement. The results of G506-IADO and G001-IADO are shown in Table 14 and Table 15 as given below. The rounding to the nearest half chip with phase data from the G506-IADO produced the smallest sum of root mean square (RMS) residuals between the four comparisons in Table 14.

Deviations of DGPS G506 – IADO Results from Benchmarks						
Time	48780	48795	48810	48825	48840	48855
Rounding + Phase						
ΔX	1.2079	-7.8762	11.4399	7.9121	-0.1740	-5.8655
ΔY	-9.4313	-13.3901	-11.4716	-26.9126	-25.0238	-4.0262
ΔZ	-22.9984	-2.6666	9.4155	0.7598	32.3287	2.2141
RMS Sum	RMS	RMS	RMS	RMS	RMS	RMS
135.7818	24.8865	15.7620	18.7382	28.0618	40.8823	7.4510
Rounding Only						
ΔX	0.3503	-1.6118	15.3006	13.3834	4.2889	-1.0521
ΔY	-15.9744	-34.7789	-24.2511	-33.3919	-30.3423	-6.9320
ΔZ	-12.8075	10.7457	18.0081	14.1586	34.6209	8.6464
RMS Sum	RMS	RMS	RMS	RMS	RMS	RMS
186.8015	20.4777	36.4368	33.8603	38.6600	46.2348	11.1319
Truncation + Phase						
ΔX	-0.3373	-7.5969	8.9264	3.5956	-0.1744	-5.8660
ΔY	12.2766	12.5211	-21.2244	-40.9535	-25.0237	-4.0261
ΔZ	-32.0308	-19.6336	4.8894	4.2808	32.3281	2.2135
RMS Sum	RMS	RMS	RMS	RMS	RMS	RMS
172.0036	34.3045	24.4943	23.5385	41.3333	40.8818	7.4511
Truncation Only						
ΔX	-1.1949	-1.3325	12.7871	9.0669	4.2884	-1.0526
ΔY	5.7336	-8.8678	-34.0039	-47.4328	-30.3422	-6.9318
ΔZ	-21.8398	-6.2212	13.4821	17.6796	34.6203	8.6458
RMS Sum	RMS	RMS	RMS	RMS	RMS	RMS
181.0671	22.6115	10.9141	38.7497	51.4261	46.2342	11.1314

Table 14: DGPS G506-IADO Results of Carrier Phases Combined with Pseudoranges

Deviations of DGPS G001 – IADO Results from Benchmarks						
Time	48780	48795	48810	48825	48840	48855
Rounding + Phase						
ΔX	6.5120	-19.4754	-4.5844	6.1275	3.7191	-8.8056
ΔY	-10.5241	-11.5716	5.8060	-10.0509	-5.4538	-20.4005
ΔZ	-2.0085	-16.0319	-6.4534	6.9703	20.2300	6.6335
RMS Sum	RMS	RMS	RMS	RMS	RMS	RMS
108.2565	12.5378	27.7527	9.8170	13.6804	21.2798	23.1889
Rounding Only						
ΔX	-4.4465	-6.8088	2.3100	5.6023	8.6432	-2.7331
ΔY	-3.3863	-10.0737	2.6461	8.9258	-26.0502	-22.1577
ΔZ	-13.1786	-2.1903	8.6327	-4.0696	21.3983	8.3499
RMS Sum	RMS	RMS	RMS	RMS	RMS	RMS
105.9246	14.3148	12.3546	9.3200	11.2968	34.8024	23.8360
Truncation + Phase						
ΔX	6.5096	-19.4768	-4.5858	6.1261	3.7185	-8.8062
ΔY	-10.5217	-11.5706	5.8070	-10.0499	-5.4536	-20.4003
ΔZ	-2.0107	-16.0326	-6.4542	6.9696	20.2293	6.6327
RMS Sum	RMS	RMS	RMS	RMS	RMS	RMS
108.2537	12.5349	27.7537	9.8187	13.6786	21.2789	23.1887
Truncation Only						
ΔX	-4.4488	-6.8103	2.3086	5.6008	8.6425	-2.7337
ΔY	-3.3839	-10.0727	2.6471	8.9268	-26.0500	-22.1575
ΔZ	-13.1808	-2.1910	8.6320	-4.0703	21.3976	8.3491
RMS Sum	RMS	RMS	RMS	RMS	RMS	RMS
105.9254	14.3170	12.3547	9.3193	11.2971	34.8016	23.8356

Table 15: DGPS G001-IADO Results of Carrier Phases Combined with Pseudoranges

However, there is very little difference in the summed RMS totals between the four methods in the G001-IADO results. Statistically, G001-IADO results are essentially alike, and the smallest RMS totals show virtually no difference between rounding only or truncation only of the half chip intervals as well as no real difference between rounding with phase data or truncating with phase data. A careful look shows that the differential navigation solutions for the second time (i.e. 48795 s) and for the fifth time (i.e., 48840 s) are nearly 20 to 30 meters off from the "true" difference for G001 and IADO, while the difference for the other times are around 10 meters from "true" GPS.

The primary cause for the large variation in the navigation solutions over such a short time is there are no satellites in the southern half of the hemisphere of coverage. For example, the satellites are generally scattered east and west of the local zenith (See Figure 13), which would indicate large ΔY residuals as shown in Table 10 through Table

12. This is also indicated from Table 1, where DOP_{north} = 2.26 was the largest component compared to DOP_{east} = 0.76 and DOP_{up} = 1.35. The secondary cause may well be due to some multipath problems that magnified the pseudorange errors, but there are not sufficient data and auxiliary resources to analyze multipath. One symptom of multipath is the variation of the navigation solution when the receivers are all stationary, but it is beyond the scope of this research to pursue this issue.

A closer look at the raw pseudoranges also shows that the data are not appropriate for single station navigation. For example, some of the G001 pseudoranges have L2 shorter than L1, which is untenable given that the free electrons in the ionosphere retard the lower frequency more than the higher frequency by the factor of the inverse square of the frequency. The other L1 and L2 pseudorange data pairs at G001 are nearly equal for each time measured, so the ionosphere has supposedly no real effect. However, the other stations do show the ionosphere effect contributes between 8 and 16 meters in the delay of the measured pseudoranges, which is more realistic for low solar spot activity and GPS measurements in daylight. But, even IADO and G506 ionosphere delays computed at time 48780 are not well correlated to elevation as shown in the following table

Satellite PRN	G506 Ionosphere Correction (m)	IADO Ionosphere Correction (m)	Local Azimuth	Local Elevation
2	-15.300	-16.664	80.79°	23.66°
8	-9.205	-8.845	-57.31°	68.73°
11	-15.283	-16.262	-48.71°	26.36°
27	-9.162	-10.388	-29.26°	74.94°
28	-8.940	-9.439	-83.38°	40.84°
31	-9.411	-10.314	59.34°	49.36°

Table 16: Station Ionosphere Corrections versus Elevations

Note that the zenith correction is near -9 meters and the correction near the horizon is about -15 to -16 meters. However, the mid elevation satellites have nearly the same ionosphere correction as PRN-27 nearest the zenith. This is unexpected, as some obliquity trend in the ionosphere should be measured in the pseudoranges. Thus, the data from the DGPS survey receivers do not work well for the single station navigation process that was expected. Although one could claim that multipath reflections skewed the data, there is no way to verify this with the current equipment and available resources.

Chapter 10 Conclusion

Dual phase measurements incorporated with the pseudorange data have the strong potential to obtain subcentimeter precision in pseudoranges. An in-depth examination of the Ashtech data reveals that the Ashtech receivers do not output the type of data appropriate for this test. First, the Ashtech dual receivers are designed for DGPS processing, which involve double and triple differencing, not actual navigation from a single receiver. The pseudoranges do not have to be precise, as long as all DGPS receivers are consistent, and common errors subtract out in the DGPS processing. Second, no measurement was possible to determine the phase angle at the point of the code boundary with these receivers. So, the assumption that the actual phase angles at the code boundary were zero is invalid. Third, there is no assurance that the code and carrier tracking loops are operating coherently with each other or that the measurements are captured at the same instant with both loops. The Ashtech internal design is proprietary, and that precludes an examination into the software architecture to determine what characteristics may exist between pseudorange and carrier phase data recording or processing. Fourth, the method given in Chapter 9 was not conclusive, since it demonstrates the final navigation solutions were about equal in error with or without carrier phase measurements.

What has been demonstrated in this dissertation is that the:

(1) New linear exact GPS navigation algorithm has been derived as an alternative to the standard iterative GPS method,

(2) Standard iterative GPS navigation solution may stall prematurely in many small regions, which are dependent on the satellite configuration, without getting to the receiver's actual location,

(3) New carrier phase method combined with pseudorange data can obtain subcentimeter precision in the GPS pseudoranges, and

(4) As demonstrated in the simulation, the combination of the exact GPS algorithm with the near perfect pseudoranges can obtain very accurate navigation solutions, even in deep space where the standard GPS navigation method will converge to a near-Earth solution, instead.

The second item may happen more frequently with the advent of pseudolites placed nearby airports for precise GPS approach and landings of aircraft. The last item allows far more precision of a GPS navigation solution that has been unattainable until now, but it requires the proper code and carrier phase measurements, especially at the code boundary, to provide the extra precision that will obtain the accurate navigation.

The advantages of combining both carrier and code measurements by this method will allow real-time calculations without the required convergence of the standard GPS techniques that may take several minutes to settle with similar precision. Furthermore, this exact navigation solution is not affected from previous signal interruptions, such as blockage or intermittent lightening strikes, which greatly affects smoothing process such as Kalman filtering. When the GPS pseudorange measurements are precise within a centimeter, the new navigation algorithm can be as accurate and undoubtedly quicker than the standard GPS iterative techniques.

There were many lessons learned from this effort. The GPS data must be appropriate for this technique to work, but a redesigned GPS receiver should be able to accomplish the necessary measurements with minimal modifications to the software architecture. Also, the simplest designs are often the hardest to verify in practice, because the means to operate under the simple conditions had not been conceived yet to make the measurements. However, the confidence to continue with sound theory is what made this research worth pursuing.

Chapter 11 Bibliography

Ashby, N., 1990, "A Tuitorial on Relativistic Effects in the Global Positioning System," NIST Report, 40RANB9B8112, Boulder, CO.

Ashby, N., August 1987, "Relativistic effects in the Global Positioning System", Proceedings of the International Association of Geodesy (AG) Symposia of the XIX General Assembly, Vol. 1, pp 41-50.

Bancroft, S., 1985, "An Algebraic Solutio of the GPS Equations." *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-21, No. 6, pp. 56-59.

Bauersima, I., 1983, GPS II, Radiointerferometrische Satellitenbeobachtungen, Mitteilungen der Satelliten-Beobachtungsstation Zimmerwald, Bern, Vol. 10, 1983.

Berman, A., 1976, "The Prediction of Zenith Refraction from Surface Mesurements of Meteorological Parameters," JPL TR-32-1602, Jet Propulsion Laboratory.

Biton, I., Koifman M., and Bar-Itzhack, I. Y, 1996, "Direct Solution of the GPS Equations," AIAA J. of guidance, Control and Dynamics, Proceedings of the ION GPS-96 Conference, Vol. 2, p. 1313-1320.

Black, H., 1978, "An Easily Implemented Algorithm for the Tropospheric Range Correction," *Journal of Geophysical Research*, Vol. 83(B4), p. 1825-1828.

Black, H., and Eisner, A., 1984,"Correcting Satellite Doppler Data for Tropospheric Effects." *Journal of Geophysical Research*, Vol. 89.

Boithias, L., 1982, *Radio Wave Propagation*, McGraw Hill, New York.

Brunner, S., and Hajj, G. A., 1991, "An Improved Model for the Dual Frequency Ionospheric Correction of GPS Observations," *Manuscripta Geodaetica*, Vol. 16, pp.205-214.

Chao, C., March 1974 "The Tropospheric Calibration Model for Mariner Mars, 1971," JPL TR 32-1587, Jet Propulsion Laboratory, Pasadena, CA.

Davis, J., et al., 1985, "Geodesy by Radio Interferometry: Effects of Atmospheric Modeling Errors on Estimates of Baseline Length," *Radio Science*, Vol. 20.

Davies, K., 1989, *Ionospheric Radio*, Peter Peregrinus, Ltd., London.

Elgered, G., Johansson, J., and Rönnäng, B., November 1985, "Methods to correct for the Tropospheric Delay in Satellite-Earth Range Measurements," in *Proceedings of the Second SATRAPE Meeting*, Saint-Mandé, France, p. 60-76.

Goad, C., and Goodman, L., Dec. 12-17, 1974, "A Modified Hopfield Tropospheric Refraction Correction Model," *American Geophysical Union Annual Fall Meeting*.

Goodman, J. M., and Aarons,J., 1990,"Ionospheric Effects on Modern Electronic Systems," *Proceedings of the IEEE*, Vol. 78, pp. 512-582.

GPS Joint Program Office (GPS JPO), 23 June 1998, SS-GPS-300E-001, System Specification for the NAVSTAR Global

GPS Joint Program Office (GPS JPO), 25 September 1997, ICD-GPS-200C, Interface Control Document of the Global Positioning System, Section 6.2.1. (Initial release 10 Oct 1993)

GPS Joint Program Office (GPS JPO), 7 December 2004, ICD-GPS-200D, Interface Control Document for Global Positioning System, Version D.

GPS Joint Program Office SMC/GP, 2006, http://gps.losangeles.af.mil/gpsoverview.htm.

Hatch, R., Feb. 8-12, 1982, "The Synergism of GPS Code and Carrier Measurements," *Proceedings of the Third International Symposium on Satellite Doppler Positioning*, (New Mexico State University), pp. 121-123.

Herring, T., 1992, in Refraction of Transatmospheric Signals in Geodesy, ed. J. de Munck & T. Spoelstra (Delft: Netherlands Geod. Commission), Vol. 157.

Hofmann-Wellenhof, B., Lichtenegger, H., and Collins, J., 1994, GPS Theory and Practice, Springer-Verlag, New York, p. 122.

Hopfield, H., April 1969,"Two Quartic Tropospheric Refractivity Profile for Correcting Satellite Data," *Journal of Geophysical Research*.

Hopfield, H., July 1970,"Tropospheric Effect on Electronmagnetically Measured Range: Prediction from Surface Weather Data," Applied Physics Laboratory, Johns Hopkins University.

Hotovec, J., 1980, *Tropospheric Delay Correction*, GPS-FSD #549, November 24, 1980.

Hwang, P. Y. C., and Brown, R. G., 1989, "GPS Navigation: Combining Pseudorange with Continuous Carrier Phase Using a Kalman Filter,: *Proceedings of ION GPS-89*, (Colorado Springs, CO), Institute of Navigation, Sept. 27-29, pp. 185-190.

Janes, H., Langley, R., Newby, S., March 1989, "A comparison of several models for the prediction of tropospheric propagation delay," Proceedings of the Fifth International Geodetic Symposium on Satellite Positioning, Las Cruces, NM, Vol. 2, p. 777-788.

Jeyapalan, K., Stein, M. A., Awuah-Baffour, R. et al, 1993, "Use of GPS for Photogrammetry," Engineering Research Institute, Iowa State University, HR-342.

Kaula, W., 1966, Theory of Satellite Geodesy, Blaisdell Publishing Company, Waltham, MA.

Klobuchar, J. A., 1996, Ionospheric Effects on GPS, Global Positioning System: Theory and Applications, Parkinson and Spilker (eds.), Vol. 1, Chapter 12, Progress in Astronautics and Aeronautics, American Institute of Aeronautics and Astronautics, Inc.

Kovach, K., January 2000, "New User Equivalent Range Error Budget for the Modernized Navstar Global Positioning System", ION National Technical Meeting, Anaheim, CA

Krause, L. O., 1987, "A Direct Solution to GPS-Type Navigation Equations," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-23, No. 2, pp. 225-232.

Lanyi, G., July 1984, "Tropospheric Calibration in Radio Interferometry," in *Proceedings of the International Symposium on Space Techniques for Geodynamics*, Sopron, Hungary, Vol. 2, p. 184-195.

Leick, A., 1995, GPS Satellite Surveying, 2nd Edition, Chapter 8, John Wiley & Sons, Inc., pp. 294-295.

Lorrain, P. and Corson, D., 1970, Electromagnetic Fields and Waves, W. H. Freeman and Company, San Francisco, p. 501 or p. 576.

Marini, J., 1972, "Correction of Satellite Tracking Data for an Arbitrary Tropospheric Profile," *Radio Science*, Vol. 7.

Marini, M., and Murray, C., 1973, "Correction of Laser Range Tracking Data for Atmospheric Refraction at Elevations Above 10 Degrees,", NASA/GSCF X-591-73-351.

Marion, J. B., 1970, Classical Dynamics of Particles and Systems, 2nd Edition, Chapter 8, Academic Press.

Mayer, M., Kutterer, H., Kargoll, B., and Heck, B., July 2000,"The Impact of Troposphere Models on GPS-derived Point Positions in Antarctica," in IAG International Symposium on Gravity, Geoid and Geodynamics, Banff, Alberta, Canada.

Millman, G. H., 1965, "Atmospheric Effects on Radio Wave Propagation," *Modern Radar, Analysis, Evaluation, and System Design*, edited by R. S. Berkowitz, John Wiley & Sons, New York.

Millman, G. H., and Reinsmith, G. M., 1974, "An Analysis of the Incoherent Scatter-Faraday Rotation Technique for Ionospheric Propagation Error Correction," General Electric Technical Information Series Report, R74EMH2.

Parkinson, B. W., 1996, "Introduction and Heritage of NAVSTAR: The Global Positioning System," Progress in Astronautics and Aeronautics, Vol. 163, Chapter 1.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., 1997, Numerical Recipes in C, 2nd edition, Cambridge University Press, 1997.

Rahnemoon, M., 1988, "Ein Korrekturmodell für Mikrowellenmessungen zu Satelliten," German Geodetic Commission, Munich, Series C, Vol. 335.

Saastamoinen, J., 1972a, 1972b, 1973, "Contribution to the Theory of Atmospheric Refraction," *Bulletin Géodésiqué*, Vol. 105, Sept. 1972, Vol. 106, Dec. 1972, Vol. 107, March 1973.

Saastamoinen, J., 1972, "Atmospheric Correction for the Troposphere and Stratosphere in Radio Ranging of Satellites," *Geophysical Monograph 15*, American Geophysical Union.

Sears, F. W., 1949, *Optics*, Addison-Wesley.

Spilker, J. J., 1995, "Digital Communications by Satellite," Prentice Hall, Englewood Cliffs, NJ.

Spilker, J. J., 1996, GPS Signal Structure and Theoretical Performance, *Global Positioning System: Theory and Applications*, Vol. 1, Progress in Astronautics and Aeronautics, Vol. 163, Chapter 3, American Institute of Aeronautics and Astronautics, Inc.

Spilker, J. J., 1996a, Tropospheric Effects on GPS, *Global Positioning System: Theory and Applications*, Vol. 1, Progress in Astronautics and Aeronautics, Vol. 163, Chapter 13, American Institute of Aeronautics and Astronautics, Inc.

Spilker, J. J., 1996b, Satellite Constellation and Geometric Dilution of Precision, *Global Positioning System: Theory and Applications*, Vol. 1, Progress in Astronautics and Aeronautics, Vol. 163, Chapter 5, American Institute of Aeronautics and Astronautics, Inc.

Yionoulis, S., 1970, "Algorithm to Compute Tropospheric Refraction Effects on Range Measurements," *Journal of Geophysical Research*, Vol. 75(36), p. 7636-7637.

Zumberge, J. F., and Bertiger, W. I., 1996, Ephemeris and Clock Navigaiton Message Accuracy, *Global Positioning System: Theory and Applications*, Vol. 1, Progress in Astronautics and Aeronautics, Vol. 163, Chapter 16, American Institute of Aeronautics and Astronautics, Inc.

Appendix: Satellite Position Calculated at Transmission

Through any two points is a unique line. Through any three points is a unique quadratic and so forth with more planar points. The interpolating polynomial of degree N-1 through N points where $y_1 = f(x_1)$, $y_2 = f(x_2)$, …, $y_N = f(x_N)$ is solved explicitly by Lagrange's interpolation formula by assuming the orbit is effectively coplanar:

$$P(x) = \frac{(x-x_2)(x-x_3)\ldots(x-x_N)}{(x_1-x_2)(x_1-x_3)\ldots(x_1-x_N)}y_1 + \frac{(x-x_1)(x-x_3)\ldots(x-x_N)}{(x_2-x_1)(x_2-x_3)\ldots(x_2-x_N)}y_2 + \ldots$$
$$+\ldots+\frac{(x-x_1)(x-x_2)\ldots(x-x_{N-1})}{(x_N-x_1)(x_N-x_2)\ldots(x_N-x_{N-1})}y_N$$

There are N terms, each having a polynomial of degree N-1 and having to be zero at all of the x_i except one, which is constructed as y_i. Lagrange's formula gives no error estimate, and has to be recalculated for each x. A better algorithm for constructing the same, unique interpolating polynomial is Neville's algorithm (Press et al, 1997).

Let P_1 be the value of the zeroth degree polynomial through the first point (x_1, y_1). This means $P_1 = y_1$. Likewise, define the other points P_2, P_3, …, P_N. Define P_{12} to be the value at x of the unique polynomial of degree one passing through both (x_1, y_1) and (x_2, y_2).

Then, define the other values P_{23}, P_{34}, …, $P_{(N-1)N}$. Continue doing this for the higher order polynomials to $P_{12\ldots N}$, which is the value of the unique polynomial through all N points to obtain the desired answer. The various P values form a table of ancestors on the left leading to a single descendant on the extreme right. For example, with N = 4,

$$\begin{array}{lllll} x_1: & y_1 = P_1 & & & \\ & & P_{12} & & \\ x_2: & y_2 = P_2 & & P_{123} & \\ & & P_{23} & & P_{1234} \\ x_3: & y_3 = P_3 & & P_{234} & \\ & & P_{34} & & \\ x_4: & y_4 = P_4 & & & \end{array}$$

Neville's algorithm is a recursive way of constructing the table one column at a time from left to right. It is based on the relationship between ancestors and descendants.

$$P_{i(i+1)\ldots(i+m)} = \frac{(x - x_{i+m})\,P_{(i+1)\ldots(i+m-1)} + (x_i + x)P_{(i+1)(i+2)\ldots(i+m)}}{x_i - x_{i+m}}$$

The recurrence works because the two ancestors already agree at points $x_{i+1} \ldots x_{i+m-1}$.

Neville's algorithm was applied to the five GPS ephemeris points for each satellite as calculated by JPL on March 19, 2002, between 13:00 to 14:00 hours Greenwich Mean Time in 15 minute intervals. Data were gathered for the six satellites of SV-2, 8, 11, 27, 28 and 31. The satellite positions were interpolated by this method at the received GPS time. This produced the initial information to calculate the pseudoranges. One iteration was then needed to obtain the transmission time interval due to the finite speed of light, which was subtracted from the reception time to get the time at transmission. Neville's algorithm was applied again to find the interpolated satellite positions at transmission time.

On the following table, the results to calculate the XYZ and satellite clock bias for SV-2 are shown. The five tiers used to set up the interpolation for Neville's algorithm are given both in symbolic terms and with numeric values. At the calculated time of transmission in seconds since the beginning of the day, the final result of SV-2's position in kilometers and satellite clock bias expressed in meters is:

(20549.69, -10293.6, 13593.08) and -169.516 meters as given in the table below.

The 5 ephemeris points are chosen from the JPL ephemerides at 15 minute intervals with the center point near the transmission time from SV-2. Each interpolation is run separately for the individual components of XYZ and the satellite clock bias.

Interpolation at Transmission Time of Satellite ID = 2				
T = 48779.9215576643				
P1	P2	P3	P4	P5
P12	P23	P34	P45	
P123	P234	P345		
P1234	P2345			
P12345				

Table 17: Interpolation by Neville's Algorithm Using SV-2

Interpolation for X coordinate at transmission time T				
16804.05	18619.83	20249.34	21658.76	22819.99
20798.62	20575.1	20531.1	20729.67	
20552.76	20548.71	20550.95		
20549.79	20549.6			
20549.69				
Interpolation for Y coordinate at transmission time T				
-11050.6	-10651.4	-10345.6	-10108.9	-9912.13
-10172.3	-10284.4	-10298.3	-10266.4	
-10295.6	-10292.7	-10295.1		
-10293.5	-10293.7			
-10293.6				
Interpolation for Z coordinate at transmission time T				
17800.12	16047.97	14026.7	11768.31	9309.286
13945.54	13622.63	13575.22	13735.74	
13590.35	13594.19	13591.27		
13593.16	13593.02			
13593.08				
Interpolation for Satellite Clock Correction at transmission time T				
-169.504	-169.509	-169.515	-169.52	-169.529
-169.516	-169.516	-169.516	-169.513	
-169.516	-169.516	-169.515		
-169.516	-169.516			
-169.516				

Table 17: Interpolation by Neville's Algorithm Using SV-2 (Continued)

Acknowledgments

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting this research. First and foremost, Dr. Kandiah Jeyapalan for his patience and support, especially his willingness to teach several classes through the Iowa Communications Network, so that I could attend classes away from campus while working full time during my program of study. His flexibility to schedule field assignments was crucial so that I could commute and personally attend the sessions to gain the experience necessary during my program of study. I also thank my fellow graduate coworkers, Dwipen Bhagawati, Panjetty-Kumaradevan Senthil Kumar, and Arumbuliyur-Comondoo Ramesh, who worked with me, both in classes and field projects to hone our surveying skills and collect the data used for this research. I deeply appreciate all the assistance provided by Dr. Lee Ann Willson as my Minor Professor so that I could complete my minor program requirements of physics under the Department of Physics and Astronomy. Most of all, I am deeply grateful to my wife, Arlette, for her words of encouragement that renewed my hopes and the continual support to write this thesis and complete my graduate education.