International Global Navigation Satellite Systems Society
IGNSS Symposium 2007

The University of New South Wales, Sydney, Australia
4 – 6 December, 2007

Recent GNSS Developments at NavCom Technology

Ron Hatch
NavCom Technology, Inc., USA
Phone 01 310 381 2603 / Fax 01 310 381 2001 / Email rhatch@navcomtech.com

Jerry Knight
NavCom Technology, Inc., USA
Phone 01 310 381 2609 / Fax 01 310 381 2001 / Email jknight@navcomtech.com

Liwen Dai
NavCom Technology, Inc., USA
Phone 01 310 381 2647 / Fax 01 310 381 2001 / Email ldai@navcomtech.com

ABSTRACT

NavCom has developed techniques that limit the effects of interfering signals, reduce the code and carrier tracking errors caused by multipath reflections and improve the reliability and accuracy of long distance RTK. Strong continuous wave interference is overcome by looking only at the sine wave peaks of the interfering signals. Pulse jamming signals are automatically detected and blanked. The detailed shape of signal phasers at code edge transitions are used to determine the amplitude and phase of the interfering multipath signals and to remove their effects on the code and carrier tracking loops. The reliability and accuracy of long distance RTK is improved by solving for distance-dependent biases and the use of a partial ambiguity search technique.

KEYWORDS: GNSS receivers, electromagnetic interference, multipath mitigation, long distance RTK, carrier phase.

1. INTRODUCTION

NavCom Technology, Inc. is a wholly owned subsidiary of Deere and Company and is part of their Intelligent Mobile Equipment Technologies (IMET) group. NavCom broadcasts high precision GNSS augmentation signals (StarFire) that enable decimetre positioning globally and designs high-accuracy, multi-frequency GNSS receivers with a special emphasis on the needs of the high precision agricultural market.

This report describes three new developments and provides results that demonstrate each improvement. A patent application has been submitted for each development.

The first new development involves a simple but dramatic way to improve the receiver performance in the presence of continuous wave or pulsed interference. It builds upon earlier work by Amoroso (1983) and significantly improves an implementation of the concept by Maenpa (1997).

Second is a new technique to improve the receiver resistance to multipath in both the carrier-phase and the pseudorange measurements. The new technique measures the behaviour of the complex carrier at code edges and compares the measurements to those from a known multipath-free signal to determine the amplitude and phase of the composite multipath signal.

Third is a set of improvements to carrier-phase differential (RTK) application software. A modified LAMBDA technique is used to allow a hierarchy of partial ambiguity resolutions and allow some distance dependent biases to be included as part of the solution. The result is both better accuracy and higher probability of correct ambiguity resolution at longer distances.

2. MITIGATION OF INTERFERENCE

2.1 Background

While not required, a four bit analog to digital (A/D) converter is assumed in the following description. The four bit A/D converter samples both the I (cosine) and the Q (sine) of the signal carrier. The A/D converters are assumed to operate at a rate that exceeds the Nyquist rate of the information content in the input signals. The maximum positive value corresponds to an A/D reading of 15 and the maximum negative to an A/D reading of zero. The zero reading corresponds to half way between an A/D reading of 7 and 8.

The A/D samples are monitored by an automatic gain control (AGC) that adjusts the receiver gain so that a desired fraction of the samples have a magnitude that is greater than a threshold. The gain-adjusted samples are then converted into 3-level, sign-magnitude data that are used by the digital processing ASIC.

2.2 Automatic Gain Control (AGC)

The AGC classifies each I or Q A/D sample to be either Active (large magnitude) or Inactive (small magnitude). See Table 1. At the start of each set of N samples, a counter is zeroed. During the following N sample interval, the counter is incremented when the magnitude of an I or Q samples exceeds the threshold. N is chosen so that the AGC input can be generated at an approximate 200 KHz rate.

A/D (binary)	Sign - Magnitude	AGC	3-Level
1111	+ 7	Active	+1
1110	+ 6	Active	+1
1101	+ 5	Active	+1
1100	+ 4	Active	+1
1011	+ 3	Active	+1
1010	+ 2	Active	+1
1001	+ 1	Active	0
1000	+ 0	Inactive	0
0111	- 0	Inactive	0
0110	- 1	Active	0
0101	- 2	Active	-1
0100	- 3	Active	-1
0011	- 4	Active	-1
0010	- 5	Active	-1
0001	- 6	Active	-1
0000	- 7	Active	-1

Table 1. A/D to AGC and 3-Level Sample Conversions

At the end of the N samples, the count is compared to a programmable limit. If the count exceeds the limit, the gain is too high so the gain is reduced. If the limit is not exceeded, the gain is too low and the gain is increased. The AGC feedback filter natural frequency is designed to be between 100 ms and 10 seconds.

2.3 Mitigation of Continuous Wave Interference

Continuous wave (CW) interference degrades the signal to noise ratio of GNSS navigation signals. The CW interference can be viewed as a sinusoidal interfering wave that overlays the spread spectrum signal used by the GNSS signal. Before the spread spectrum signal is correlated, the amplitude of the CW signal is often significantly larger than the amplitude of the spread spectrum signal. Correlation despreads the energy of the signal and spreads the energy of the interfering CW signal, which then becomes noise-like. If the additional noise from the spread CW signal is larger than the ambient background thermal noise, the signal to noise ratio of the received GNSS signal is decreased.

The signal to noise ratio varies across various portions of the interfering CW sine wave, Amoroso (1983). Figure 1 shows the combination of a noisy spread-spectrum signal and a down-converted CW jamming signal. The noise masks the underlying coded signal. Figure 2 superimposes the coded GNSS signal and also displays the AGC quantization thresholds on the curve from Figure 1. The rate of change of the CW sine wave at its crest and trough is nearly zero, so the imposed spread spectrum code is more observable. The rate of change is largest on the slopes of the sin wave, and the imposed spread spectrum code is difficult to discern.

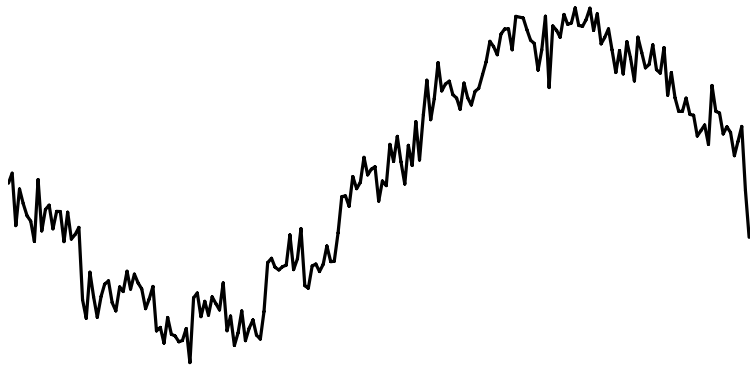

Figure 1. Noisy CW-Jammed Signal

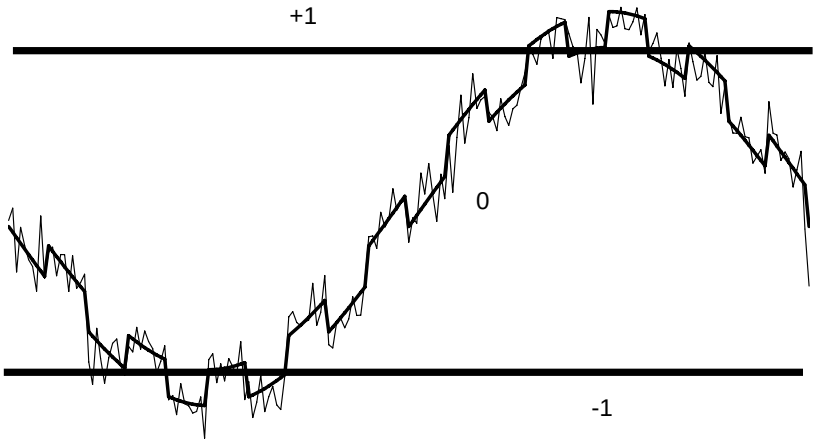

Figure 2. Amaroso Sampling of Jammed Signal

Amoroso found the receiver's performance is greatly enhanced in the presence of CW jamming if only the samples that fall near the crests and troughs of the CW sine wave are used during signal processing. The 10 to 20% of the samples that fall at the crest are weighted +1 (Figure 3), the 10 to 20% that fall at the trough are weighted -1, and the remaining samples are discarded by giving them a weight of 0.

Amoroso achieved the desired sample populations by carefully controlling the amplitude of the signals going into the A/D converters so that only the very crests of the CW waves are sampled. Implementation of a similar control mechanism has proved to be difficult, but possible, in GPS receivers, Maenpa (1997). The problem is that the amplitude of the noise in GNSS A/D samples is large compared to the amplitude of the spread spectrum signal. Thus, a small error in the estimation of the noise amplitude causes the A/D converters to pass too much or too little of the spread spectrum code.

This shortcoming can be overcome by controlling the sample statistics rather than adjusting the signal amplitude and by taking advantage of some fortuitous statistical properties of Gaussian noise and sinusoidal jamming signals. The A/D samples in GNSS receivers are noise limited, i.e. the amplitude of the noise is much greater than the amplitude of the signal. And the noise is known to have normal, i.e. Gaussian, statistics.

For a sample population with normal, Gaussian statistics, a third of all samples have magnitudes within 0.43 standard deviations of the sample mean (Figure 3). A third of all samples are greater than the mean plus 0.43 standard deviations and the remaining third of all samples are less than the mean minus 0.43 standard deviations.

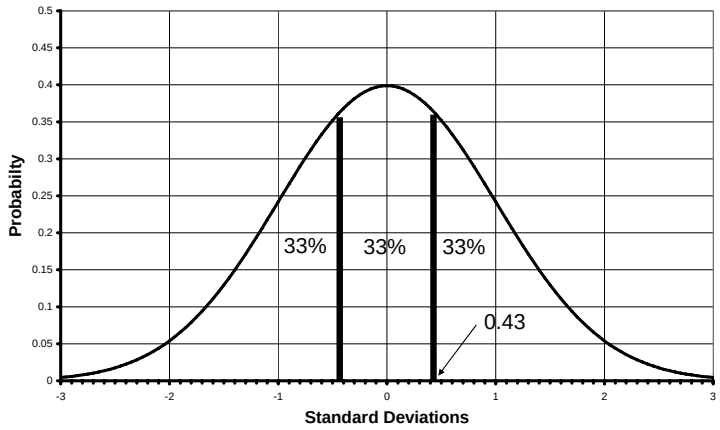

Figure 3. Population Distribution for AGC

For Gaussian noise with a uniformly distributed phase approximately 60% of all samples occur within 0.86 standard deviations (twice previous) of the mean (Figure 3). In this case, 20% of the samples are larger than the mean plus 0.86 standard deviations and 20% are less than the mean minus 0.86 standard deviations.

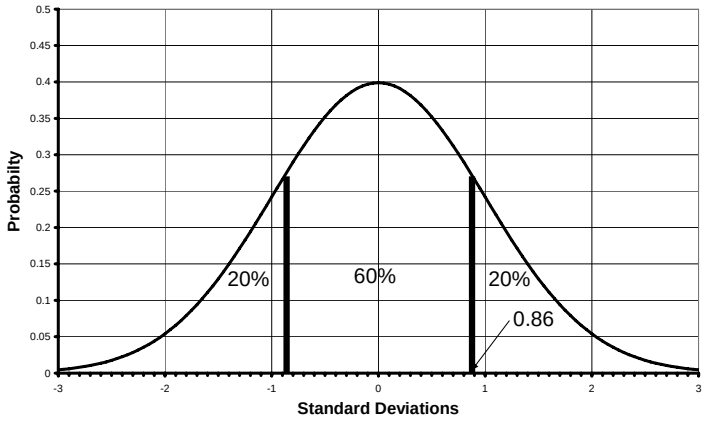

Figure 4. Population Distribution for 3-Level Samples

For a CW signal that is larger than the Gaussian noise, the probability that $|\cos(\theta)| > 0.5$ is

0.667 since cos(60º) = 0.5. If the AGC is configured so that 0.667 of all A/D samples have a large AGC magnitude (active), an A/D magnitude of twice the AGC activity corresponds to a cosine of 1.0, which is the crest of the CW wave, and is exactly the sampling that is required by Amoroso when a strong CW jammer is present. Therefore, if we configure the AGC threshold to provide 2/3 active samples and then only use the samples that have a magnitude greater than or equal to twice the AGC threshold for signal processing, we get near optimum performance for both unjammed and CW jammed conditions.

The AGC feedback in the receiver can easily produce a sample population with the desired features. The first step is to set the AGC activity count to 0.667 so that two thirds of the 4-bit sign/magnitude samples have a magnitude greater than or equal one (active) and one third of the samples are inactive. The second step is to configure the A/D to 3-level conversion vector as shown in Table 1 such that the 3-level samples have the statistics shown in Figure 4 for unjammed conditions and ride the crest of the jammer when strong CW jamming is present.

Using the quantization of Table 1 with the associated AGC achieves near-optimum performance for both CW-jammed and unjammed conditions. The very small difference between the performance of this method and a theoretical optimum quantization is shown in Figure 5.

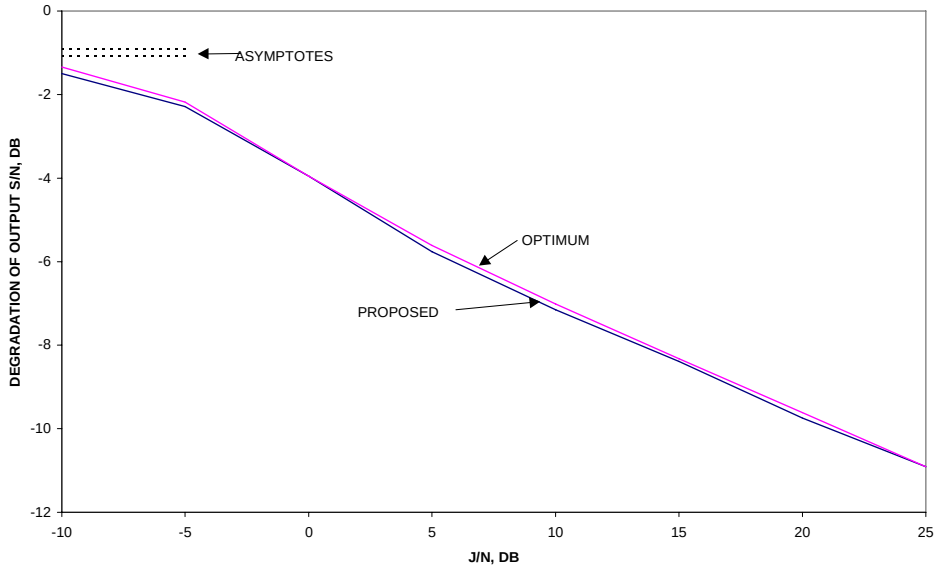

Figure 5. Comparison of Proposed and Optimum CW Jamming Performance

2.4 Mitigation of Pulse Jamming

A further mechanism is desirable to mitigate the effects of pulse jamming. If the receiver could automatically stop processing input when the pulsed jammer is on and resume processing when the jamming is off, the signal degradation is just the duty factor loss, –0.92 dB for a duty factor of 0.1.

The magnitude of the input samples can be monitored to detect a sudden increase in the proportion of very large samples caused by a strong jammer. If such a sudden increase

occurs, the receiver will stop processing the signals, i.e. enables blanking. Blanking is turned off (disabled) when the receiver detects an absence of large magnitude samples. While blanking is enabled, AGC and zero-adjust feedback are disabled.

The 3-bit magnitude of the I and Q samples are compared to a software adjustable threshold. If the magnitude of either I or Q is larger than the threshold on 12 out of 16 sample periods, the signal blanking is enabled. Once blanking is enabled, it remains enabled until at least 100 of 128 samples have both I and Q with magnitude less than the threshold.

The probability and number of standard deviations for exceeding the possible 3-bit sample magnitudes is shown in Table 2. This table assumes the desired 0.667 activity for AGC control. The probability of large samples is small, but not prohibitively small. If we wish to have a very low probability of blanking due to random noise, we must monitor a group of samples and turn blanking on only when a considerable fraction have large magnitude.

Magnitude	# Standard Deviations	Probability
1	0.43	0.666
2	0.86	0.390
3	1.29	0.197
4	1.72	0.085
5	2.15	0.032
6	2.58	0.0099
7	3.01	0.0026

Table 2. Probability of Sample of a Given Magnitude

A large number of simulations were run to find the best method for detecting pulse jamming and enabling blanking and, once blanking is enabled, for detecting the absence of pulse jamming and turning off blanking. The simulations demonstrated that blanking should be enabled if the magnitude of either I or Q is greater than or equal to four for 12 of 16 sample periods. The simulations demonstrated that there should be some hysteresis between turning blanking on and off, so that in the presence of a weak pulse jammer blanking is not toggled on and off at a high rate. Good performance is obtained if blanking is turned off when both I and Q are less than four for 100 of 128 sample periods.

An external blanking signal is also available which enables blanking whenever it is asserted. This feature may be used when a cooperative interfering signal is being broadcast, such as a radio transmission or a cooperative pseudolite. Automatic blanking can be disabled by setting the threshold to 0x7, the largest possible sample magnitude. Blanking is typically disabled for the first few seconds after the receiver is turned on to allow the AGC and zero balance functions to reach steady state.

3.0 Multipath Mitigation

3.1 Background

Global satellite navigation receivers (GNSS), such as GPS, GLONAS or GALILEO, use range measurements that are based upon line-of-sight signals from the satellites. The Navigator measures the time-of-arrival of the broadcast signal. This time-of-arrival

measurement includes a time measurement based upon the coded portion of the signal, called pseudorange, and phase measurements based upon the L-band carrier of the signal. Ideally, these measurements are based only upon the direct line-of-sight signal. However, the actual signals received by user equipment are a composite of the direct line-of-sight signal and at least one, more likely many, secondary reflected signals. These secondary signals are reflected by any number of structures, including buildings, equipment, even the ground. The strength of these reflected, or multipath, signals is based upon the efficiency of the reflector (the reflectivity) and the added path length. An illustration of this composite signal is provided in Figure 6 for a single reflected signal. In the case illustrated, A_d is the direct path signal and A_m is the reflected (multipath) signal. The multipath signal, necessarily, has a path that is longer than the direct path signal. Since the path is longer, the multipath signal is a replica of the direct path signal only slightly delayed by the added path length and lower amplitude.

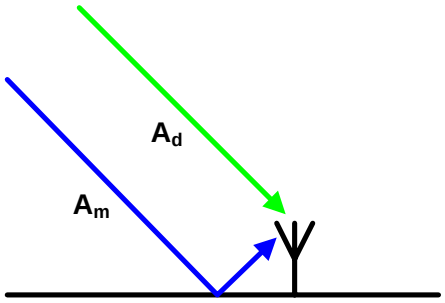

Figure 6a. Composite Signal Model

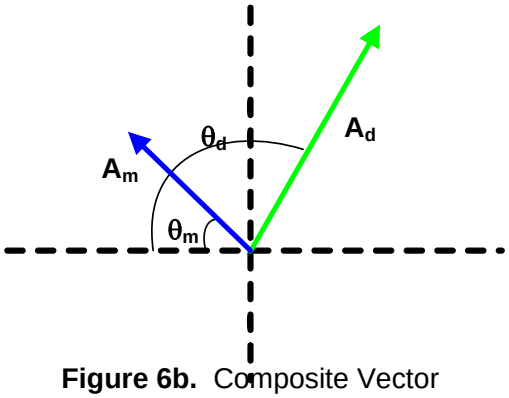

Figure 6b. Composite Vector Model

Phase Multipath is the distortion in the L-band carrier caused by reflected signals. Figure 6b illustrates the signal received by the GNSS equipment in vector form; it is comprised of the direct path signal vector and the multipath signal vector. The direct path signal has amplitude, A_d, and phase relative to the receiver's internal reference, θ_d. The multipath signal arrives at a different time so it has a different phase, relative to the receiver's internal reference, θ_m, and different, lower, amplitude A_m.

The signal received by the GNSS receiver is shown in Figure 7 in sin-wave form. The undistorted direct carrier wave from the satellite is represented by the medium-amplitude (medium weighted) curve. In this diagram the multipath signal, the lowest amplitude curve, is delayed by 45 degrees and has relative amplitude of 0.5. The GNSS receiver tracks the composite of the direct and the multipath waves, which is shown as the largest amplitude, heaviest curve. A GNSS code chip edge occurs in Figure 7 when the phase of the direct signal is 360 degrees. The GNSS codes are bi-phase modulated, so they are encoded by advancing or retarding the carrier phase by 90 degrees. The reflected signal experiences the same phase shift, but it is delayed relative to the direct signal by 45 degrees. The zero crossings of the composite signal in Figure 7 are delayed (shifted right) from those of the direct signal. If the path length difference resulted in a phase shift of the multipath signal of greater than 180 degrees, it would appear to cause a phase advance of the composite signal. This apparent phase advance or delay produces the phase tracking errors caused by the multipath interference.

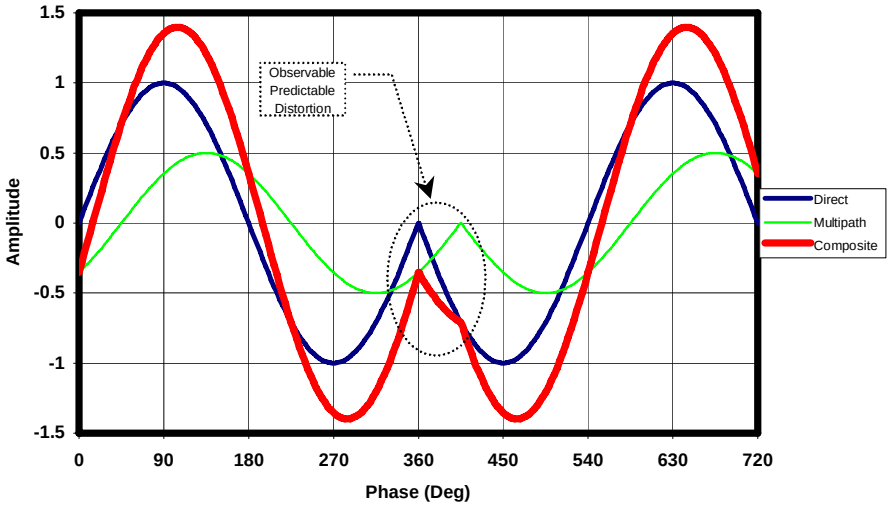

Figure 7. Carrier Phase Code Edge Transition

The composite wave in the figure has a very complex, but predictable, shape in the interval between the direct and reflected bit edges (360 to 405 degrees in the diagram). The new technique described uses the observable characteristics of the received composite signal in this interval to determine and remove the multipath-caused phase tracking errors.

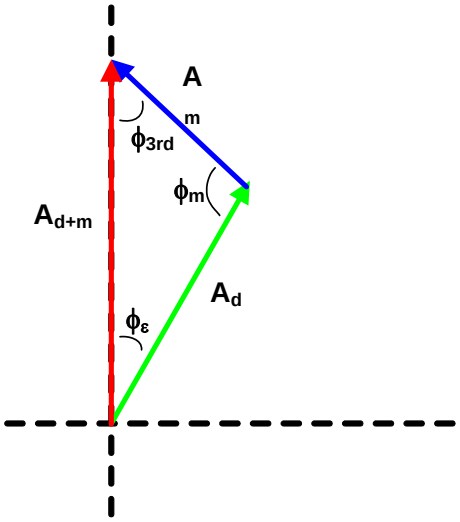

Figure 8. Tracked Vector Model

The basic signal processing performed by the GNSS receiver on the composite signal is a tracking loop process that matches the phase of a replica signal created by the receiver with the phase of the signal received from the satellite. The timings used to create the replica signal provide the basic code pseudorange and carrier phase measurements made by the receiver. The tracking process of the receiver cannot distinguish between the direct path signal and the multipath signal, so it tracks the composite signal, which is shown in vector form in Figure 8. The composite signal, A_{d+m}, is the vector sum of the direct path signal, A_d, and the multipath signal, A_m. The signal replica has a phase error of ϕ_ε relative to the direct path signal.

3.2 The New Phase Multipath Mitigation

At each code edge, the I and Q components of the direct and multipath signal transition between normalized amplitudes of +1 and -1. The transitions are filtered by the satellites and the receiver. GPS satellite transitions behave approximately like a 6-pole, approximately 24 to 30 MHz bandpass Butterworth filter.

The response of a 6-pole Butterworth Filter with an IF equivalent bandwidth of 30 MHz is shown in Figure 9. The normalized signal amplitude goes through zero at about 140 nanoseconds in the figure and starts a damped oscillation around -1 at about 160 nanoseconds. 30 MH is wider than most current high precision GNSS equipment but bandwidths of this magnitude are becoming more standard as signal processing speeds increase. The step response of this filter requires almost 50 ns to transition the phase reversal, and steady state is not attained for almost 150 ns. The delays are greater for narrower filters. Figure 9 illustrates the direct signal and a multipath signal with relative amplitude of 0.25 and a delay of 15 ns. The composite of the direct and multipath signals is also shown.

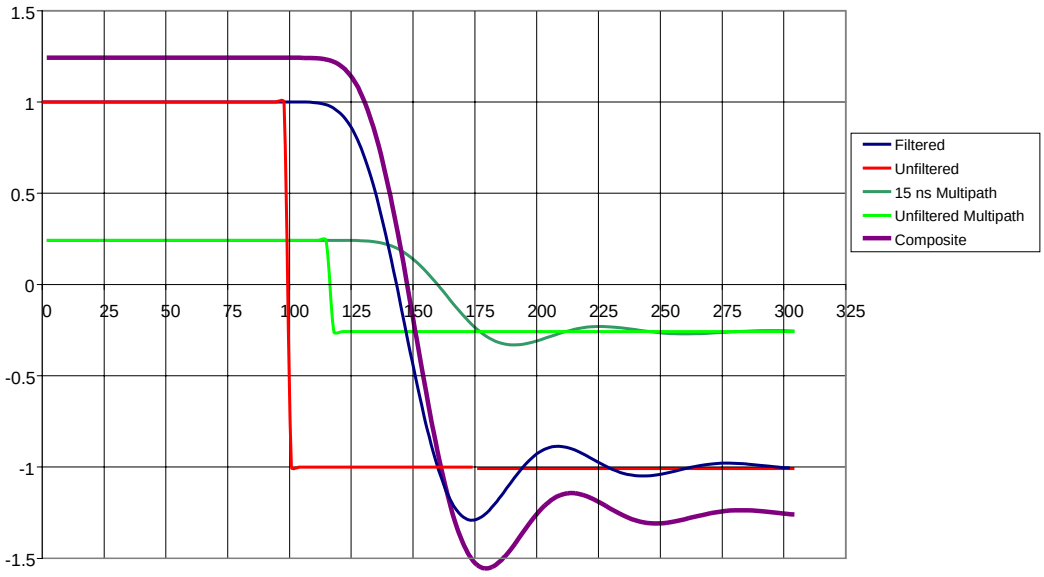

Figure 9. MultiPath Step Response (nanoseconds)

Where path length differences are less than the step response time, the transition is a trajectory (Figure 10), where the multipath signal begins to transition before the line-of-sight signal has concluded its transition.

The input satellite signal samples can be categorized according to the phase of the code or carrier NCO. Classically, the carrier tracking loops use all of the samples, because this provides the best signal-to-noise advantage. It has long been recognized that for best code-multipath rejection only the portion of the code edges that are near the chip transitions (for example, the Magnavox MX4200 implemented +/- 0.25 chip narrow correlation starting in 1988). This type of narrow correlator is functionally equivalent to summing only samples that are made when the code phase fraction is between 0.75 to 0.25 chips (0.75 to 1.0 and 0.0

to 0.25). Samples with phase between 0.25 and 0.75 (> 0.25 and < 0.75) are discarded and not used in the processing.

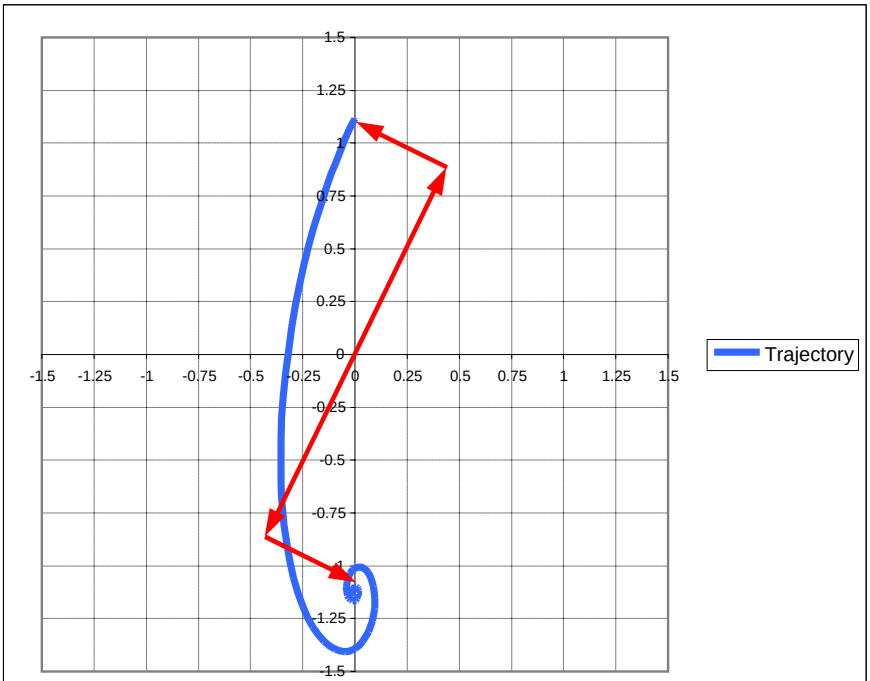

Figure 10. Transition Trajectory

The samples needed to trace out the transition curve of Figure 10 are produced by selectively integrating samples in exactly the same way. The I and Q sums are only enabled when the phase of the replica coder is within the bounds required to produce the desired sample intervals covering the transition path.

Classically, receivers have categorized their sample integrations by sample number relative to the code edge. The receivers typically make a nearly exact-integer number of samples per chip and enable the accumulations by the sample number. For example, if the receiver makes four I-, Q-sample pairs per chip, it achieves a plus or minus one-quarter chip correlation by numbering the samples 1 through 4 relative to the code edge and by summing two subsets of the samples. The first subset only includes the sample 4, the one that occurs just before the code edge. The second subset only includes the sample 1, the one that occurs just after the code edge.

The fine sample intervals needed to cover the phase transition curve of Figure 10 may be 5 ns or less in size. A complex sampling and data processing rate of at least 200 MHz is required if the sampling numbering method is used to create such narrow sampling subsets. Such a high rate is very expensive in terms of power consumption, parts cost and implementation difficulty, but does provide the best possible signal to noise ratio because there is at least one sample for every interval from every code transition.

An alternative method can be used to achieve the same results with much slower sampling rates; the sample rate may be as low as 30 MHz, the lowest speed allowed by the Nyquist

limit. This method uses a sampling rate that is not an integer multiple of the code chipping rate. Therefore, the timing of the samples with respect to the code edge changes every code edge. All that is required to provide fine-spaced samples is a correspondingly fine test of the replica code phase. Since the spacing of the input samples in time is larger than the desired time span for each of the subsets, only a fraction of the subsets are assigned a sample each code edge. However, over the span of many chips, every subset receives many samples. This method does have a relative disadvantage with respect to signal to noise ratio because only a fraction of the subsets receive a sample each code edge; but, since the time correlation of multipath is very long, the integrations can be continued for longer periods in order to achieve the required signal to noise ratio. Integration times of 100 ms to several seconds are sufficient and are much shorter than the correlation period of short range multipath.

Classically, high precision GNSS receivers have only provided a few correlator sums per channel; for example punctual, early and late – and often early and late are combined in the hardware and output as a single I- and Q- pair. Some GNSS receivers that have been optimized for very fast start up time have provided many I- and Q-accumulations during signal acquisition, but these receivers typically revert to using only 2 or 3 of the accumulations per channel when they are tracking the satellite signals.

Ten to twenty accumulations per channel are sufficient to track the signal trajectory depicted in Figure 10. As a bonus, during signal acquisition the additional accumulators can be used to increase the code search rate.

If the exact shape of the trajectory of Figure 10 can be observed, the effect of the multipath distortions can be determined and removed. As shown in figure 9, multipath interference has several observable effects upon the composite signal. Among these are: The amplitude of the composite signal is changed. The time between the start and end of the filtered bit reversal is increased. The time between the start of the bit reversal and the zero crossing is increased. As shown in figure 9 the amplitude, shape and zero crossing of the composite curve are a direct function of the multipath interference.

The basic concept of the estimation technique is that the transition trajectory and errors of Figure 10 can be modelled by observing the deviations between the observed composite signal and the computed direct signal. The required observations are simply the two dimensional I and Q samples obtained during the transition region shown in Figure 10. They provide both an amplitude and phase of the composite signal. The extent by which the path of the transition deviates from the expected path with no multipath allows one to compute the amplitude, phase and delay of the multipath signal and thereby remove it. In so doing, the tracking error, Φ_ε, represented in Figure 8, induced by the multipath can be robustly estimated. The algorithm that estimates this tracking error is based upon the model of the transition trajectory.

The transition trajectory of Figure 10 can be modeled by the set of equations below. Referring to the tracked signal model of Figure 8, the goal of the technique is the estimation of the parameters that define the trajectory of Figure 101.

$$\delta \equiv \text{Multipath Delay}$$
$$\phi_\varepsilon \equiv \text{Tracking Error}$$
$$SR(t) \equiv \text{IF Filter equivalent LowPass Step Response}$$

$$I(t) = A_d \bullet SR(t) \bullet \cos(\phi_\varepsilon) + A_m \bullet SR(t+\delta) \bullet \cos(\phi_{3rd})$$
$$Q(t) = A_d \bullet SR(t) \bullet \sin(\phi_\varepsilon) - A_m \bullet SR(t+\delta) \bullet \sin(\phi_{3rd})$$

$$\phi_{3rd} = \pi - (\phi_\varepsilon + \phi_m) \qquad \text{3rd angle of multipath triangle}$$

$$I(t) = A_d \bullet SR(t) \bullet \cos(\phi_\varepsilon) + A_m \bullet SR(t+\delta) \bullet \cos(\pi - (\phi_\varepsilon + \varphi_m))$$
$$Q(t) = A_d \bullet SR(t) \bullet \sin(\phi_\varepsilon) - A_m \bullet SR(t+\delta) \bullet \sin(\pi - (\phi_\varepsilon + \phi_m))$$

$$\left.\begin{aligned} I(t) &= A_d \bullet SR(t) \bullet \cos(\phi_\varepsilon) - A_m \bullet SR(t+\delta) \bullet \cos(\phi_\varepsilon + \phi_m) \\ Q(t) &= A_d \bullet SR(t) \bullet \sin(\phi_\varepsilon) - A_m \bullet SR(t+\delta) \bullet \sin(\phi_\varepsilon + \phi_m) \end{aligned}\right\rangle$$

The algorithm will estimate the following parameters:

- A_d The amplitude of the direct path signal
- A_m The amplitude of the multipath signal
- ϕ_ε The phase tracking error induced by the multipath
- ϕ_m The multipath angle relative to the direct path signal
- δ The delay of the multipath signal relative to the direct path signal

The algorithm will require as inputs the following parameters:

- SR(t) The step response of the IF filter expressed as a Low Pass equivalent
- A_{d+m} The amplitude of the composite signal derived from the steady state signal tracking loop
- I(t),Q(t) The I and Q baseband measurements taken at multiple time delays during the Code transition

These equations are non-linear with respect to the unknown quantities and there is no closed form solution. Each of the code tracking correlators provides a measurement of the amplitude of the I and Q components of the composite signal vector at the corresponding correlation times. Thus there are many more measurements than unknown quantities. A combination of a least-squares or Kalman filter with an iterative, non-linear root finding technique is suitable for solving the equations.

4.0 Long Distance RTK

4.1 Background

High precision GPS kinematic positioning in real time is increasingly used for many surveying, navigation and machine guidance applications on land, at sea and in the air. Standard practice is for one reference receiver to be located at a station whose coordinates are known, while the second receiver's coordinates are determined relative to this reference

receiver. For the highest accuracy, carrier phase measurements are used. However, the use of carrier phase data has a price in system complexity because the measurements are ambiguous at the whole cycle level. Therefore, an ambiguity resolution (AR) process must be included as an integral part of the processing software. Real time kinematic (RTK) systems are used for many time-critical high precision applications.

There are three general classes of techniques used to resolve the unknown integer ambiguities. These are: (1) search techniques in the measurement domain (orbit free techniques); (2) search techniques in the coordinate domain (particularly suitable when good approximate initial coordinates are available); and (3) search techniques in the estimated ambiguity domain using least-squares or Kalman filter estimation. The search can be implemented for the original ambiguity parameters or for combinations of the original parameters. In the former case, the search is made of the integer ambiguities directly, using techniques such as: the Fast Ambiguity Resolution Approach (FARA) (Frei & Beutler, 1990); the Least Squares Ambiguity Search Technique (LSAST) (Hatch, 1990); the Cholesky decomposition method (Euler & Landau, 1992); the Fast Ambiguity Search Filter (FASF) (Chen, 1993). In the latter case, the Least-square AMBiguity Decorrelation Adjustment (LAMBDA) method performs the search using transformed integer ambiguities (Teunissen, 1994). The transformed integer ambiguities are constructed from the original ambiguities and can be used to recover the original integer ambiguities. The advantage of this procedure is that the transformed real-valued ambiguities have smaller standard deviations. Consequently, the computation time is significantly reduced. However, all of the methods have difficulty in fixing ambiguities to the correct integer values if the floating ambiguities are significantly biased by multipath, un-modeled systemic errors or cycle slips.

When GPS signals are continuously tracked and no loss-of-lock occurs, the integer ambiguities resolved at the beginning of a survey can be kept for the entire GPS kinematic positioning span. However, the GPS satellite signals may be occasionally shaded (e.g. due to buildings in "urban canyon" environments) or momentarily blocked (e.g. when the receiver passes under a bridge or through a tunnel). Generally, in such cases the integer ambiguity values are "lost" and must be re-determined. This process can take from a few tens of seconds to several minutes. In fact, the presence of significant multipath errors or unmodeled systematic biases in one or more measurements of either pseudorange or carrier phase may make it impossible with present commercial GPS RTK systems to resolve the ambiguities. As the receiver separation increases, the distance-dependent biases (e.g. orbit errors and ionospheric and tropospheric effects) grow and, as a consequence, reliable ambiguity resolution (or re-initialization) becomes an even greater challenge. When the integer ambiguities are resolved improperly, it can be very difficult to identify which ambiguity is incorrect. But it is clear that one or more of the estimated floating ambiguities must be off by some number of whole cycles.

At NavCom we have developed a new generation of Long Range RTK software by directly tackling the problems of distance dependent biases and developing techniques to eliminate or mitigate the choice of improperly resolving the integer cycles. The techniques employed are described below. Following the description of the techniques employed, sample results are given.

4.2 A Partial Search Technique to Avoid Erroneous Whole Cycle Determination

For rapid ambiguity resolution, a minimum of five satellites is required. When six or more

satellites are observed, some of the ambiguities can be arbitrarily removed from the search process. By computing all combinations of five or more satellites, it may be possible to isolate one or more of the floating ambiguities with significant biases, and a successful ambiguity resolution may result. This procedure could be implemented in the software by eliminating one (or more) satellites (keeping at least five). If the ambiguity resolution fails, the procedure can be repeated until the resolution is successful. If all possible sets of five or more satellites have been tested and the ambiguity test still fails, the ambiguity resolution procedure has failed at least until more measurement data is collected. The above procedure of searching through subsets significantly increases the ambiguity resolution success rate. However, these combined search processes take significant time to compute, and it is generally impossible to search all combinations in the software embedded within the GPS receivers because of limited processing resources.

A partial search technology is proposed which takes advantage of information about the floating ambiguities to select those partial ambiguity combinations most likely to yield a successful search. The resulting partial search technique makes GPS RTK fix availability much higher, especially for long-range applications and in challenging environments. In addition, the technique does not require substantial computational power and, as a result, is suitable for GPS receivers with limited microprocessor capabilities.

The partial search technique described below assumes that either a least-squares or Kalman filter have been used to arrive at an estimated vector, $\hat{X}_N$, of real-valued ambiguities and an associated variance-covariance matrix, $P_{\hat{X}_N}$, of those ambiguities. Specifically, it involves a modification of LAMBDA technique mentioned above. The key to the LAMBDA method is the computation of the transformation matrix for constructing the multi-satellite ambiguity combinations. X.W Zhang (2005) presented a Modified LAMBDA method for integer least-squares estimation. MLAMBDA includes a modified Z matrix reduction process and modified search process. The ambiguity transformation matrix Z reformulates the original ambiguity vector as the transformed ambiguity vector, whose variance-covariance matrix has much smaller diagonal elements.

$$\hat{Z}_N = Z \cdot \hat{X}_N \tag{1}$$

$$P_{\hat{Z}_N} = Z P_{\hat{X}_N} Z^T \tag{2}$$

The goal is to find the vector of integer ambiguities, N_k^Z, which minimizes the value of R_k where:

$$R_k = (\hat{Z}_N - N_k^Z)^T P_{\hat{Z}_N}^{-1} (\hat{Z}_N - N_k^Z) \tag{3}$$

The original ambiguity estimation problem has been transformed. The new problem is to search for a new integer set that makes and passes the validation and rejection criteria tests.

In order to ensure that the transformed ambiguity has integer characteristics, the transformation matrix Z has to have integer entries. In order to ensure that the original ambiguity can be determined from the transformed ambiguity, the inverse of the transformation matrix also has to have integer entries. Therefore, the matrix Z is an

admissible ambiguity transformation if and only if the matrix Z has integer entries and its determinant equals 1. After obtaining a vector of integer ambiguities that satisfy the validation criteria, the original ambiguities are recovered via:

$$N_k = Z^{-1} N_k^Z \tag{4}$$

After the double-difference measurements from both frequencies are processed, the floating ambiguity vector and the associated variance can be obtained from either a sequential least-squares implementation or from a Kalman filter estimation. Although the L1 and L2 floating ambiguities are the parameters to be estimated, the ambiguity pairs to be searched in LAMBDA method, can be different. In the technique here described, $\hat{N}_1$ and the wide lane $\hat{N}_{(1,-1)}$ are chosen for the search process. They are formed from the original L1 and L2 float ambiguity vector of equation (5) using the following equation:

$$\begin{pmatrix} \hat{N}_1 \\ \hat{N}_{(1,-1)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \hat{N}_1 \\ \hat{N}_2 \end{pmatrix} \tag{5}$$

From equation (2), the variance-covariance matrix of $\hat{N}_1$ and the wide lane $\hat{N}_{(1,-1)}$ is obtained as follows:

$$\mathrm{P}_{\hat{N}_1, \hat{N}_{(1,-1)}} = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} P_{N_1} & P_{N_1 N_2} \\ P_{N_2 N_1} & P_{N_2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} \tag{6}$$

The best integer vector candidate and second best integer vector can be found during the Lambda search process. If the ambiguity validation discrimination tests have been passed, the best integer ambiguity vector will be accepted as the correct ambiguity set to produce the fixed position solution and its associated variance-covariance matrix. If no ambiguity pairs are excluded from the search process, the search results of $\hat{N}_1, \hat{N}_{(1,-1)}$ and $\hat{N}_1, \hat{N}_2$ are identical. However, they will differ if some ambiguities are excluded from the search process. It is well known that the wide lane ambiguity is easy to resolve due to the longer, 0.86 cm, wavelength. The common practice is to first fix the wide lane ambiguity followed by the L1 ambiguity. The partial fix procedure is applied when the combination fix of $\hat{N}_1, \hat{N}_{(1,-1)}$ fails to pass the discrimination test criterion. The partial search technique is to exclude those satellites whose integer ambiguities are different between the best candidate set and the second best candidate set from the LAMBDA search process. In most cases, one or more $\hat{N}_1$ ambiguities are excluded in the partial search, but all of the wide lane ambiguities remain in the search. This partial search process can be repeatedly applied until the minimum of five ambiguities no longer remain to be searched. If the partial $\hat{N}_1, \hat{N}_{(1,-1)}$ search fails, only $\hat{N}_{(1,-1)}$ or partial $\hat{N}_{(1-,1)}$ ambiguities can be used in the LAMBDA search process. It is obvious that the partial fix procedure can be used to increase the fix availability, especially in long range applications where biases are more common.

If the fix or partial fix successfully passes the ambiguity discrimination test, equations (7-8) can be used to recover the original L1 and L2 ambiguities states and the variance- covariance

in the Kalman filter.

$$\begin{pmatrix} \hat{N}_1 \\ \hat{N}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} \hat{X}_{N_1} \\ \hat{X}_{N_{(1,-1)}} \end{pmatrix} \tag{7}$$

$$\mathrm{P}_{\hat{N}_1,\hat{N}_2} = \begin{pmatrix} 1 & 0 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} P_{N_1} & P_{N_1 N_{(-1,1)}} \\ P_{N_{(-1,1)} N_1} & P_{N_{(-1,1)}} \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix} \tag{8}$$

Note, if $\hat{N}_1$ is floating but $\hat{N}_{(1,-1)}$ is fixed, the fractional part of $\hat{N}_1$ and $\hat{N}_2$ is exactly the same, and the variance of Q_{N_1}, Q_{N_2} and covariance $\mathrm{Q}_{N_1 N_2}$ are the also the same. If $\hat{N}_1$ is fixed later, the fixed value of $\hat{N}_1$ and $\hat{N}_2$ will be adjusted to an integer value and the variance and co-variance will become zero.

The partial search procedure described above is easily modified to handle the triple frequency situation which will arise in the modernized GPS and in the future European Galileo system. After the floating ambiguity and variance-covariance of the three original phase measurements (L1, L2 and L5) are estimated in the Kalman filter, the original ambiguity states at L1, L2 and L5 are transformed into the three alternate ambiguity combinations, $\mathrm{N}_{(1,-1,0)}$, $\mathrm{N}_{(0,1,-1)}$ and $\mathrm{N}_{(1,0,0)}$. The rest of the procedure is simply an expanded implementation of the two frequency equations above.

The proposed partial search approach has significant advantages: (1) it is a rigorous approach from a Kalman filtering implementation point of view; (2) all other approaches to fix ambiguity and wide lane ambiguities, e.g. L1/L2 search and wide lane only searches, are simply special cases of the above process, and; (3) it can significantly improve RTK ambiguity resolution availability and reliability, especially for long-range applications and in other challenging environments.

4.3 Distance Dependent Bias Mitigation

In addition to the partial search described above, NavCom has added several techniques to the process aimed at improving the accuracy and increasing the range at which the RTK system can be used. As the receiver separation increases, the problem of accounting for distance-dependent biases grows and, as a consequence, reliable ambiguity resolution becomes an even greater challenge. The major challenge is that the residual biases or errors after double-differencing can only be neglected for ambiguity resolution purposes when the distance between the two receivers is less than about 10 to 15 kilometers. For longer distances the distance-dependent errors, such as orbital error and ionospheric and tropospheric delays, become significant problems. Determining how long the observation span should be to obtain reliable ambiguity resolution is a challenge for real-time GPS kinematic positioning. The longer the observation span that is required, the longer the "dead" time during which precise positioning is not possible. The ambiguity resolution process is required at the start of GPS navigation and/or surveying and whenever too many of the GPS signals are blocked or attenuated such that cycle slips or measurement interruptions occur. Quality control is critical and is necessary during all processes: data collection, data processing and data transmission. The quality control and validation criterion for ambiguity resolution represents a significant challenge.

There are three major sources of distant dependent biases. They are ionospheric and tropospheric refraction effects on the signal transmission and errors in the transmitted GPS orbits. A Kalman filter algorithm is used to estimate the ionospheric and tropospheric delays via functional and stochastic models. Orbital corrections generated from the NavCom/Deere StarFire™ system are used to limit the orbital errors. Together, the estimation of the atmospheric error sources and the use of more precise orbits, limits the distance dependent biases and results in GPS real-time kinematic (RTK) performance that is significantly better than any current products in the market, especially for long-range applications.

The standard RTK Kalman states, position, velocity, (perhaps) acceleration, and whole cycle ambiguities will not be addressed here. Instead, only the new unusual states, i.e. the residual differential tropospheric refraction and residual differential ionospheric refraction states, are discussed. Also addressed is a means for minimizing the orbital error and a way to make the ambiguity resolution less sensitive to the distant dependent biases.

4.3.1 Residual differential tropospheric refraction state

GPS satellite signals are refracted by the lower neutral part of the earth's surface extending from the surface up to 16 km in height, which is known as the troposphere, and which is composed of dry gases and water vapor. The troposphere is a non-dispersive medium and therefore the refraction does not depend upon the GNSS signal frequencies. The magnitude of the tropospheric error depends upon the satellite elevation angle and is equal to about 2.3 m in the zenith direction. It increases to over 25 m for an elevation angle of 5 degrees. The troposphere error can be largely compensated using different models such as the UNB, Hopfield or Saastamoinen model, etc. The dry component can be modeled with high accuracy, but the smaller wet component is much more difficult to model. After applying a model, the differential tropospheric error, mainly the wet component, varies typically from about 0.2 to 0.4 parts per million (ppm) of the distance separation between the base and rover receiver. The tropospheric bias is generally represented with a model for the zenith delay together with a mapping function to obtain the delay at any given satellite elevation angle. The spatial and temporal characteristics of the residual tropospheric delay can be characterized by probabilistic laws or statistical models. The effects of the troposphere on radio wave propagation then can be predicted over varying spatial dimensions and temporal scales according to a given probability density function or stochastically in terms of the spatial and temporal correlations of the fluctuations. In general, the residual tropospheric delay can be considered a first-order Gauss-Markov process.

All the deviations of the atmospheric conditions from standard conditions are subsumed within a scaling factor for the zenith delay. After the tropospheric delay model is applied, residual double differential tropospheric delay candelay can be approximated by equation (9) as a function of Residual Tropospheric Zenith Delay (RTZD) scale factor and a mapping function with respect to the elevation angle:

$$RTZD = T_{kl}^{pq} / [MF(\varepsilon^p) - MF(\varepsilon^q)] \qquad (9)$$

Where ε^p and ε^q are the average elevation angles of the two receivers for satellites p and q respectively. No matter the elevation of a satellite, the $\nabla\Delta T$ will be scaled by the map function factor of the rover location. This is obviously an approximation. Empirically, the residual zenith tropospheric delay is modeled as a first-order Gauss-Markov process. The transition matrix and dynamic model is:

$$\phi_{k,k-1} = e^{-\beta_{trop}(t_k - t_{k-1})} \tag{10}$$

$$Q_k = \frac{\sigma^2_{trop}}{2\beta_{trop}}(1 - e^{-2\beta_{trop}(t_k - t_{k-1})}) \tag{11}$$

where $1/\beta_{trop}$ is the correlation time of the troposphere wet component and σ^2_{trop} represents the wet component variance. Typically both are functions of the baseline length and height differences between base and rover receivers.

4.3.2 Residual differential ionosphere refraction states

The ionosphere is that region of the earth's atmosphere in which solar radiation causes atoms to ionize such that free electrons exist in sufficient quantities to significantly affect the propagation of radio waves. The height at which the ionosphere starts is about 50 km and extends to heights of 1000 km or more. The ionosphere is a dispersive medium for radio waves, that is, its refractive index is a function of the frequency. The ionosphere advances the carrier phase, thus causing the carrier phase measurements to be decreased; but delays the code modulation, thus causing the pseudorange measurements to be increased. The magnitude of the induced ionospheric error varies proportional to the number of free electrons, which is directly dependent upon solar radiation effects. It is, therefore, different for daytime and nighttime and from one season to another. Diurnally, the ionospheric error usually reaches a first peak at 14:00 local time, a second peak at 22:00 local time and drops to a minimum just before sunrise. Under extreme conditions, the ionospheric error can reach 15 m in the zenith direction and more than 200 m at elevations near the horizon. The ionosphere is typically the largest error source for differential processing and varies from one part per million (ppm) of the baseline length during low ionospheric periods at mid latitudes to greater than 10 ppm at low geomagnetic latitudes during midday. The parameters of the Klobuchar empirical model are broadcast by the GPS satellites in real time and provide a means to remove about 50% of the ionospheric refraction effects upon the measurements. The broadcast model is used to reduce the induced position error for single-frequency users and is also widely used to minimize the effects of the ionosphere upon the dual frequency carrier phase measurements used in high-precision RTK implementations.

After the measurements are adjusted by the broadcast ionospheric model and differenced with the reference site measurements, the remaining ionospheric effects are estimated in a Kalman filter as an element of the state vector. The dynamics of the residual error are modeled by a first order Gauss-Markov process. The transition matrix $\phi_{k-1,k}$ and dynamic model Q_k are given by:

$$\phi_{k-1,k} = e^{-\beta_{ion}(t_k - t_{k-1})} \tag{12}$$

$$Q_k = \frac{\sigma^2_{sion}}{2\beta_{ion}}(1 - e^{-2\beta_{ion}(t_k - t_{k-1})}) \tag{13}$$

$$\sigma_{sion} = \frac{\sigma_{vion} \times L}{\sqrt{1 - \left(\frac{R}{R+H} \cos(E) \right)^2}} \quad (14)$$

where $1/\beta_{ion}$ is the correlation time of differential ionosphere bias, typically from 30 to 300 seconds. σ_{sion} and σ_{vion} represent the variance of the differential slant and vertical ionosphere biases. σ_{vion} is a function of the local time and ionosphere activity. It will typically vary between 0.5ppm-2ppm of the baseline separation distance between the base and rover receivers. L is that baseline length. E is satellite elevation. The height of the ionospheric layer (H) is assumed to be 350 km and the mean radius of the earth R is 6371 km.

It should be noted that unlike residual tropospheric bias, the residual ionosphere parameters are estimated for every satellite other than the reference satellite. The ionospheric correction from global or regional ionosphere models such as WAAS can be considered as virtual measurements and incorporated into the Kalman filter estimation. However, the accuracy of the broadcast and WAAS models is insufficient to estimate the integer ambiguities in carrier phase-based applications. Typically the ground station network is too sparse or limited in geographic area to model more than the large-scale structure of the ionospheric electron content. As a result, the small to medium-scale structure of the ionospheric electron content is too inaccurate to allow direct carrier-phase ambiguity resolution.

The concept of using reference station networks for real-time kinematic (RTK) GPS positioning has recently been promoted strongly by both academic institutions and GPS manufacturers. The 'correction terms' for the atmospheric biases and orbit errors generated from network reference stations can be used to mitigate the distance dependent errors. However, the performance of these network RTK implementations depends on the linearity of the corrections with distance. This means that the non-linear portion of the distant dependent error remains and is not removed in the typical network RTK implementation. The improved modeling of the residual ionosphere, as described above, can significantly increase the performance and reliability of the ambiguity resolution process and reduce the time to first RTK fix for both long range RTK and for network RTK. Network RTK ionospheric corrections could be considered as virtual measurements in current Kalman Filter as

$$\hat{X}_{iono} = \nabla \Delta \hat{I} \quad (15)$$

Residual non-linear ionospheric biases can be estimated and the improved modeling of the residual ionosphere can increase fixing performance and reliability for network RTK.

4.3.3 Satellite orbital error mitigation from WADGPS or GDGPS

Satellite orbit error arises from the inability to accurately predict the future position of the satellite. The broadcast navigation message includes Keplerian orbital elements and time derivatives for those elements. It is generated using the measurements from five GPS monitor stations and is updated once every two hours. Tests have shown that the orbit error typically varies between 2 to 5 meters. For long range RTK, the orbital error can become the primary error source limiting ambiguity resolution after the tropospheric and ionospheric biases have been estimated in the Kalman filter. Several regional or global differential GPS systems are available which supply measurements or measurement corrections that can be used in a navigation receiver to obtain an accuracy in the 10 centimeter region. John Deere has developed the StarFire™ system which transmits corrections via communication satellites

with a global DGPS correction data stream. The global orbital correction stream can be used to mitigate orbital error for long range carrier phase differential applications and is of significant benefit for long range RTK ambiguity resolution.

4.3.4 Floating ambiguity states

In theory, the ambiguities are constant. However, the float ambiguities which are estimated from the Kalman filter contain multipath errors and other un-modeled systematic biases such as orbital errors, residual tropospheric and ionospheric bias remaining after their estimation, etc. Based on these considerations, the ambiguities, before being fixed, are modeled as a random walk with very small dynamics.

$$\phi_{k-1,k} = 1 \tag{16}$$

$$Q_k = \delta_{amb}^2 (t_k - t_{k-1}) \tag{17}$$

where δ_{amb} is a small dynamic noise value such as 0.001 cycle. Once the integer ambiguities are fixed, they are modeled as constants and the corresponding driving noise is set to zero. In some special cases, a small dynamic noise helps to resolve some issues such as a "long time to fix." Such situations may be caused by bad data suffered from multipath at the beginning of the process. For example, when the user receiver moves from severe shading environments to open sky, the measurements may be corrupted and a long time to fix might result. On the other hand, in the typical environment, the ambiguity can be resolved in a single epoch and the small dynamic noise does not impact adversely on the time to first fix.

4.3.5 Long range RTK results

The partial search technique and solution for distance dependent biases has been successfully implemented in NavCom's GPS receivers. From a review of public material the approach described appears to be unique and innovative. Extensive testing has yielded unprecedented performance particularly at long baseline ranges. Three graphs showing the statistics of the performance as a function of baseline separation distance between the reference site and the user are given below.

In Figure 11 the sample accuracy as a function of the baseline distance is given. The typical horizontal accuracy is shown to be less than one centimeter plus 0.5 parts per million of the distance. The typical vertical accuracy is less than two centimeters and one part per million of the separation distance.

Figure 12 gives the statistics on the time to first fix at the different separation distances. As shown, for distances of less than 10 kilometers the time to first fix is not more than a few seconds. Out to 50 kilometers the probability of resolving the ambiguities within a couple of minutes is greater than 95%. At 75 kilometers there is only about a 75% probability of resolving all the ambiguities and it can take as long as 4 minutes to reach that level of probability.

Figure 13 shows that out to about 45 kilometers the reliability of correctly determining the ambiguity values is greater than 99.9%.

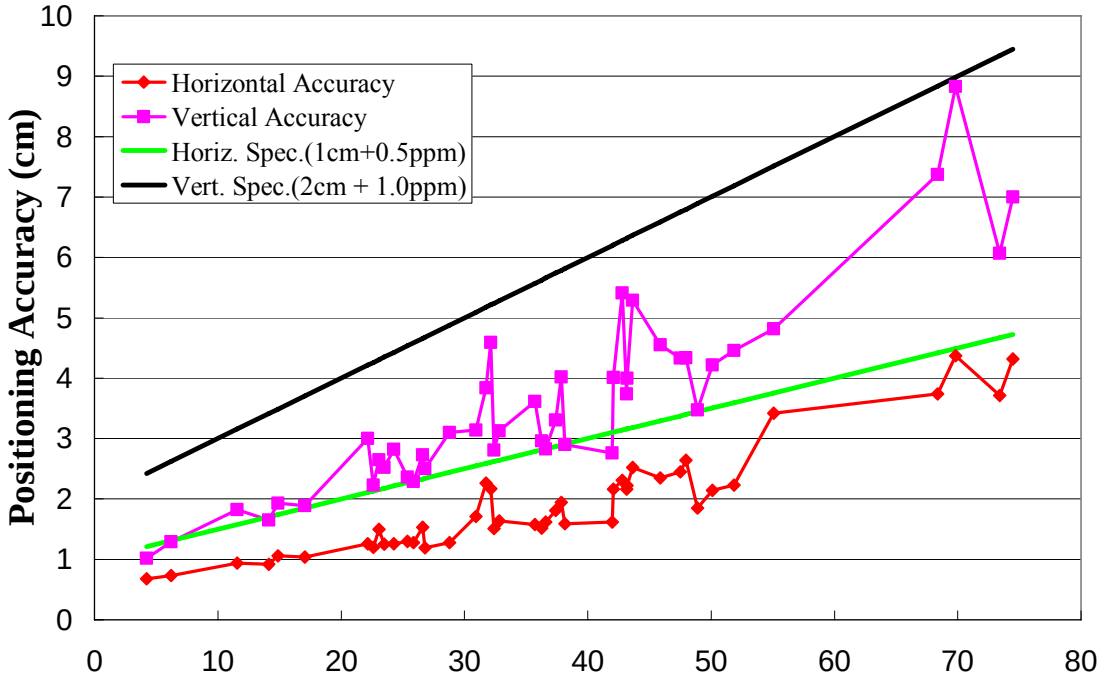

Figure 11. Baseline Length (Km)

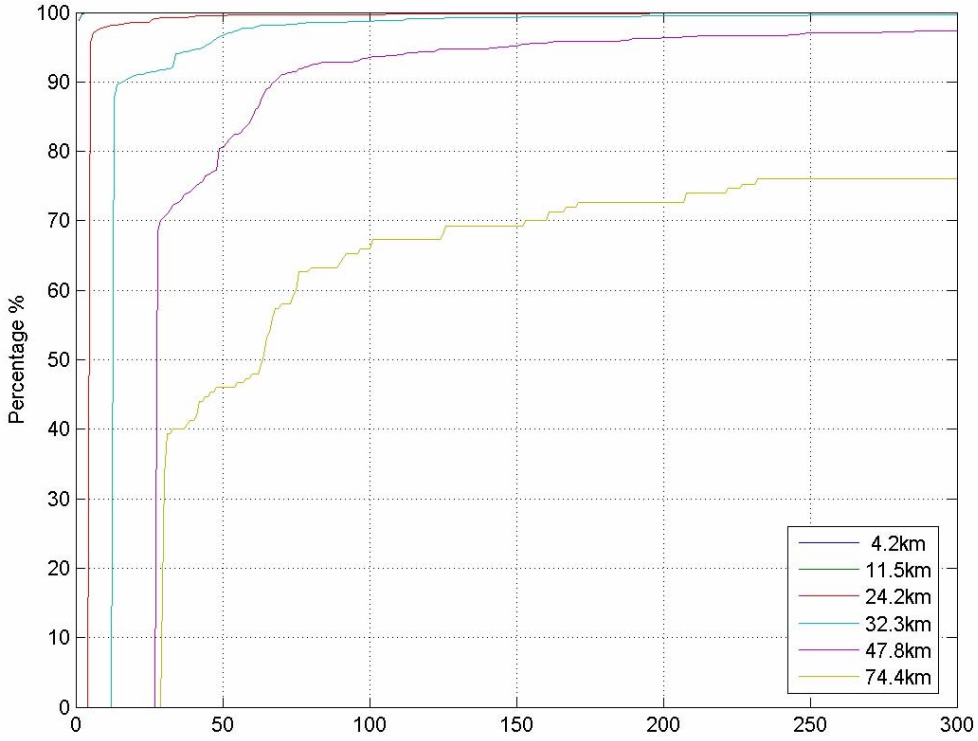

Figure 12. Time to First Fix

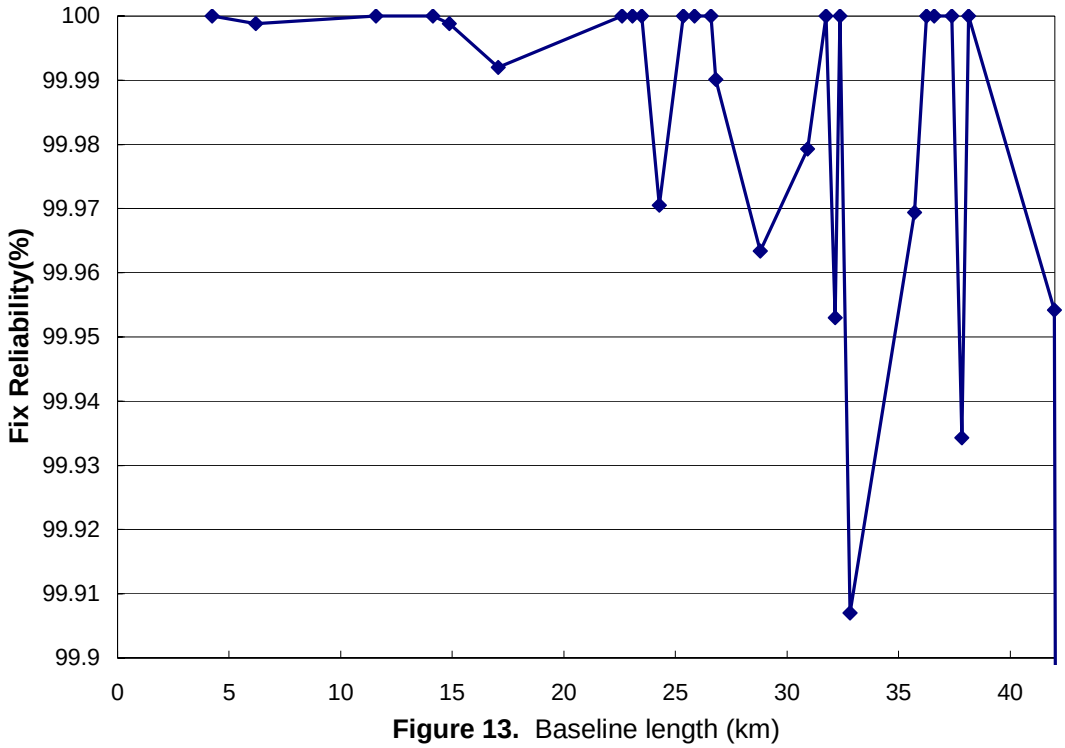

Figure 13. Baseline length (km)

5. CONCLUSIONS

Three separate developments at NavCom have been described. The first two involve new techniques employed in receiver design. The first of these two describes a simple but very effective way of mitigating interference, especially continuous wave interference. The second receiver innovation describes a new method of mitigating the effect of multipath, particularly of near-in carrier phase multipath. Finally, a new method of improving the accuracy and usefulness of long-distance RTK differential navigation is described. It makes use of a partial search technique and of solving for residual biases which are a function of separation distance to achieve a new level of accuracy and reliability at distances exceeding 10 to 20 kilometers. This new RTK capability has already been incorporated into NavCom Technology receivers.

REFERENCES

Amoroso, Frank (1983) Adaptive A/D Converter to Suppress CW Interference in DSPN Spread-Spectrum Communications, *IEEE Transactions on Communications*, Com-32(10), 1117- 1123.

Maenpa, Jon E, et. al. (1997) New Inference Rejection Technology from Leica, *Proceedings of the Satellite Navigation Section, Institute of Navigation*, September, 36-56.

Chen, D.S. (1993) Fast Ambiguity Search Filter (FASF) a Novel Concept for GPS Ambiguity Resolution, *Proceedings of ION GPS-93*, Salt Lake City, The Institute of Navigation, Washington, D.C.

Euler, H.-J. and H. Landau (1992) Fast GPS ambiguity resolution on-the-fly for real-time application, *Proceedings of Sixth International Geodetic Symposium on Satellite Positioning,* Columbus, Ohio, March 17-20, 650-659.

Frei, E. and G. Beutler (1990) Rapid Static Positioning Based on the Fast Ambiguity Resolution Approach FARA: Theory and First Results. *Manuscripta Geodaetica*, 15, 325-356.

Hatch, R. R. (1990). Instantaneous Ambiguity Resolution. *Proc. of the IAG Int. Symp. No. 107 on Kinematic Systems in Geodesy, Surveying, and Remote Sensing,* Springer-Verlag, 299-308.

Chang X.-W., X Yang, and T Zhou (2005) MLAMBDA: A modified LAMBDA method for integer least-squares estimation. *Journal of Geodesy,* 79, 552–565.

Teunissen, P. J. G. (1994) A New Method for Fast Carrier Phase Ambiguity Estimation. *Proc. IEEE Position Location and Navigation Symp. PLANS94,* Las Vegas, April 11-15, 562-573.